OPEN ACCESS

University of BRISTOL

# This electronic thesis or dissertation has been downloaded from Explore Bristol Research, http://research-information.bristol.ac.uk

*Author:*
**Maschio, Laurence**

*Title:*
**Structural and Functional Interrogation of the Abyssomicin C Biosynthetic Pathway**

**General rights**
Access to the thesis is subject to the Creative Commons Attribution - NonCommercial-No Derivatives 4.0 International Public License. A copy of this may be found at https://creativecommons.org/licenses/by-nc-nd/4.0/legalcode This license sets out your rights and the restrictions that apply to your access to the thesis so it is important you read this before proceeding.

**Take down policy**
Some pages of this thesis may have been removed for copyright restrictions prior to having it been deposited in Explore Bristol Research. However, if you have discovered material within the thesis that you consider to be unlawful e.g. breaches of copyright (either yours or that of a third party) or any other law, including but not limited to those relating to patent, trademark, confidentiality, data protection, obscenity, defamation, libel, then please contact collections-metadata@bristol.ac.uk and include the following information in your message:

• Your contact details
• Bibliographic details for the item, including a URL
• An outline nature of the complaint

Your claim will be investigated and, where appropriate, the item in question will be removed from public view as soon as possible.

# Structural and Functional Interrogation of the Abyssomicin C Biosynthetic Pathway

Laurence Maschio

A Dissertation Submitted to the University of Bristol in Accordance with the Requirements for Award of the degree of Doctor of Philosophy in the Faculty of Science

School of Biochemistry

May 2020

Word Count: 61,908

# Abstract

Natural product biosynthetic pathways produce a plethora of biologically active compounds, including antibiotics. These secondary metabolites are produced by a number of enzyme families found within bacteria, fungi and plants. These enzymes can be large, multidomain, complexes harbouring an intricate series of precisely controlled structural and functional interactions. Once their chemical products are liberated, a wide variety of tailoring enzymes perform complex and often understudied chemical reactions to create the final, active, biomolecule. This study aims to further our understanding of these processes through the characterisation of a number of enzymes within the abyssomicin C biosynthetic pathway. Type I polyketide synthases act in a concerted manner to create a linear chain, which is functionalised before cyclisation by the *bona fide* natural Diels-Alderase, AbyU, eventually leading to the potent antibiotic, abyssomicin C. A structural characterisation of the modular polyketide synthase AbyB3 *via* X-ray crystallography and cryo-EM methods aims to reconcile existing models for polyketide synthase architecture and mechanism, whilst an in-depth mechanistic characterisation of AbyU represents the first of its kind with respect to these deceptively complex catalysts.

# Acknowledgements

I would firstly like to thank my family. Thank you to my father for his support, encouragement and advice throughout my scientific education. Thank you to my mother for her support and love throughout, and especially the laundry and meal services I've been fortunate enough to receive over the past few months. Thank you also to my sister and brother for their support and patience over the past four years. A huge thank you to Ayan for making my life over the past two years so much fun. Your support through the business end of my PhD and write up has been unwavering and made my life as easy as possible.

A huge thank you to my supervisor, Paul Race, for all his encouragement and advice. My PhD wouldn't have been nearly so enjoyable or successful without your guidance. Thank you also to Alice Parnell. Without your friendship, help and advice the AbyB3 project would never have got off the ground and I would be a far worse scientist and person. Thank you to all the members of C101 who have, over the years, become amazing friends and made my PhD the fun and enjoyable experience that it was! Special mention to Claire Noble for becoming a fantastic house mate, to Jonnie Jenkins for putting up with me all the way from undergraduate labs and to Henry Stennett and Sam Williams for their company at two very enjoyable conferences. Thank you to Dora and Carl for their company at three extremely long and tiring sessions at Diamond collecting high-pressure data. To George Hutchins, Rosie Maddock, Richard Stenner and Catherine Back for all their company during various pub trips and advice in the lab, as well as to the Tuesday night football guys. Thanks too to all those who were with me in oxford for my training year and have remained such good friends – Max, Vittorio, Will, Danny and Jack.

Thank you also to Chris Willis and her lab in Chemistry. None of the AbyU work would have been possible without your synthetic chemistry magic, particular thanks to Sbu who made me huge amounts of substrate for stopped-flow experiments.

# Declaration

I declare that the work in this dissertation was carried out in accordance with the requirements of the University's *Regulations and Code of Practice for Research Degree Programmes* and that it has not been submitted for any other academic award. Except where indicated by specific reference in the text, the work is the candidate's own work. Work done in collaboration with, or with the assistance of, others, is indicated as such. Any views expressed in the dissertation are those of the author.

SIGNED: Laurence Maschio DATE: 02 May 2020

# Table of Contents

# List of Figures

# List of Tables

# Chapter 1

# Introduction

## 1.1 Antibiotics

Since the first demonstration of the antibacterial activity of penicillin by Sir Alexander Fleming in 1928,[1] antibiotics have grown to become a cornerstone of modern medicine. They are an irreplaceable tool in our fight against infectious disease and are used globally to treat bacterial infections in both human and animal populations. They are also used extensively as a prophylactic, for example in livestock, where they used to be fed year-round as common practice to mitigate the spread of infection.[2] Within human populations they are employed during and following surgical procedures, and are prescribed for individuals within high risk infection groups, including those who are immunosuppressed or are undergoing treatment for cancer. Their prevalence in developed medicine is such that almost 30 % of patients in Europe, either within a hospital or at the GP, are in receipt of a course of antibiotics at any one time.[3] This figure rises sharply within groups who are more susceptible to infection, such as the elderly and infants. In the developing world, antibiotics are often not as widely available to the local population, despite their effectiveness. This lack of availability has a huge effect on the health of the local population as well as on mortality rates and their causes. In conjunction with a number of other factors including hygiene, living conditions and lower vaccination rates, a lack of access to antimicrobial agents drives infection based mortality, and increases risks associated with routine medical procedures such as childbirth. In western society individuals now have limited memory of a time when infectious disease was a leading cause of death, whereas in Africa, infectious diseases account for 5 of the top 6 causes of death and 36.6% of all deaths in total. In South East Asia 3 of the top six causes of death are attributable to infectious diseases, and 14.3% of total deaths. In Europe, where antibiotics are freely accessible, this number falls to 2.7% of deaths.[4] These figures serve to highlight the crucial role of antibiotics in the healthcare system. Following their wider adoption at the end of the Second World War, average

life expectancy has increased by more than 65 %, from 47 years to 79 years at birth.[5] Our length and quality of life, along with economic prosperity all correlate with antibiotic accessibility.

### 1.1.1 Antibiotic Resistance

Our reliance upon antibiotics has always been tempered by the fact that bacteria develop resistance to these molecules. This is not a new phenomenon, and has been well documented since the first clinical use of antibiotics in 1940.[6] In this sense antibiotics are much like most therapeutic agents, in that if they are unable to cure the patient completely, they are only successful until resistance or tolerance has developed. This frequently occurs within oncology, as well as during virus or parasite infection. Although antimicrobial resistance (AMR) has existed for as long as antibiotics have been in use, it has grown in severity and prominence recently. The discovery of multi-drug resistant (MDR) and extensively-drug resistant (XDR) strains of bacteria[7] have made headlines and brought attention to this emerging public health crisis. The emergence of AMR has been classified as a major global public health concern by the World Health Organisation (WHO), the US based Centre for Disease Control (CDC) and the European Centre for Disease Prevention and Control (ECDC). This recognition has largely occurred as a result of both the current and future health and economic burdens imposed by AMR. In the European Union (EU) there were 400,000 drug-resistant infections recorded in 2007, which led to 25,000 deaths and an estimated €1.5 billion cost to the economy.[8] Similarly, the CDC reports 23,000 annual AMR related deaths in the USA. Predictions of the future consequences of AMR paint a bleak picture, with a 2014 UK review predicting 10 million deaths by 2050, and an economic impact of up to $10 trillion worldwide, if nothing is done to limit the prevalence of AMR.[9]

### 1.1.2 Antibiotic Modes of Action

The proliferation of AMR has led to a significant investment of time, money and expertise into the identification and development of new antimicrobial drugs. This has involved scientists, both in academia and industry, working to understand both the mechanisms of action of candidate molecules and the process of resistance emergence. On a molecular level, antibiotics typically target enzymes or proteins that perform critical functions in nucleic acid or protein synthesis, or in the construction of the cell wall. For example, β-lactams inhibit bacterial cell

wall biosynthesis, quinolones interfere with nucleic acid synthesis and tetracyclines inhibit protein synthesis.[10] Alternatively, some classes of antibiotics, such as antimicrobial peptides, create pores in the bacterial membrane, initiating lysis.[11] The most attractive targets for antibacterial intervention are those that are implicitly required for bacterial survival, with inhibition resulting in either cell death or severely impeding bacterial growth.

### 1.1.3 Antibiotic Resistance Mechanisms

It is now known that resistance significantly predates the human exploitation of antibiotics, and as such is a phenomenon that should always be of concern whenever antibiotics are used. Evolutionary studies have shown that the β-lactamase-encoding *bla*$_{\text{OXA}}$ genes are in fact several million years old.[12] Similarly a sequencing study of bacteria isolated from the Antarctic permafrost identified > 180 resistance genes spread across twenty three families of bacteria, accounting for every known mode of antibiotic resistance.[13] Resistance genes were developed and conserved in order to avoid attack by other microorganisms, and to mitigate issues of self-toxicity, brought about through the biosynthesis of bioactive molecules designed to target other bacteria and thus providing a competitive advantage.

Bacteria use a number of different strategies to negate the potency of antibiotics.[14] These fall into four key categories, each of which has been outlined in the literature. The first method employed involves making modifications to the antibiotic itself. A prominent example of this is the expression of β-lactamase enzymes, which cleave a key amide bond within β-lactam antibiotics, thereby opening the β-lactam ring essential for activity (Figure 1.1).[15] The second mechanism of resistance hinges on modifying the antibiotic target binding site. This can occur directly through the acquisition of point mutations[16] or *via post*-translational modifications such as methylation.[17] Alternatively the target binding site may be indirectly protected by mechanisms such as the imposition of steric hinderence.[18] Bacteria are also known to replace the individual targets, including through the acquisition of genes encoding redundant enzymes capable of performing the role of the enzyme targeted by the antibiotic.[19] Thirdly, bacteria can take measures to reduce the size of the cytoplasmic antibiotic pool. Bacteria have been shown to prevent internalisation access by reducing membrane permeability,[20] or, alternatively, export compounds following cell entry. This is usually achieved by efflux pumps, which are complex bacterial proteins that can eject toxic compounds in both a highly efficient and specific manner.[21] Finally, some bacteria are capable of instigating global changes throughout their cells to

evade some potent antibiotics, including vancomycin and daptomycin. Although these wide-ranging responses are not yet fully understood, they involve synergistic mutations in multiple enzymes and regulatory elements governing a whole process, such as cell membrane composition, synthesis and homeostasis.[14]

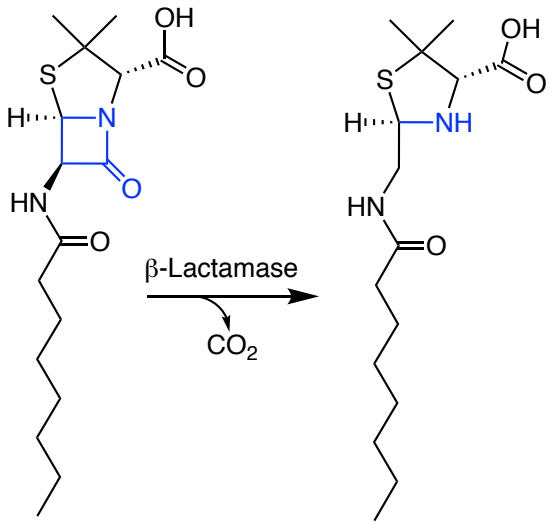



**Figure 1.1 Penicillin Inactivation by β-Lactamases.** The β-lactam ring is hydrolysed preventing interaction with the bacterial cell wall.

Crucially, populations of bacteria are not restricted to the acquisition of antibiotic resistance solely through internal processes. They are able to acquire resistance elements *via* horizontal gene transfer, a process by which genetic material is actively transferred between individual bacterial cells. This process expedites the spread of resistance, and enables MDR strains to emerge with greater frequency, particularly within closed systems, such as the gastrointestinal tract or in the hospital environment.[22]

### 1.1.4 Combatting Antimicrobial Resistance

A number of strategies have consequently been developed to counter both the spread and the impact of AMR. These include attempts to increase the size and diversity of the antibiotic

portfolio, and efforts to limit antibiotic use to only those cases where they are explicitly required. This later approach is founded on the principle that a key factor in the spread of AMR is the exposure of bacteria to low and frequent, or high and infrequent, doses of antibiotics. This provides enough of a selection pressure to drive mutation, horizontal gene transfer and the proliferation of those individual bacterial cells which are better able to withstand individual antibiotics, without killing the population in its entirety. There have been multiple efforts to reduce unnecessary antibiotic prescriptions and to reduce non-compliance with the prescribed dose. Whilst these efforts can drastically slow the spread of AMR, they will not lead to its eradication. Consequently, there is a pressing need to establish as diverse an arsenal of clinically viable antibiotics as is possible, comprising molecules with a panoply of different modes of action. Since the 1940s there have been a number of strategies employed to develop new antibiotics, which can be broadly divided into two general approaches.

### 1.1.5 Antibiotic Discovery

The first strategy is that pursued by the early pioneers of antibiotic discovery including Fleming, Duchesne and Lister. Duchesne reported the inhibition of *E. coli* by *Penicillium glaucam* a full 30 years before Fleming and became the first person to isolate the active compound. Similarly, Lister, 28 years prior to Duchesne, demonstrated the cure of a nurse's infected flesh wound using *P. glaucam* extract, after noticing that it inhibited bacterial growth.[23] This discovery route involves identifying strains of microorganisms which are able to inhibit the growth of others. The active compound(s) are then isolated, and their effectiveness determined. Compounds isolated *via* this route are subsequently tested extensively and optimised for activity against their biological target. This method of drug discovery has been by far the most prolific, yielding 20 different classes of antibiotic between the 1940s and 1970s.[24]

The second route to antibiotic drug discovery was largely enabled by advances in synthetic chemistry and seeks to develop *de novo* antibiotics able to engage 'druggable' molecular targets. Targets are usually a single protein or enzyme with a key function in an essential bacterial process required for survival. A prerequisite of this approach is that identified targets must be validated and well characterised, both *in vivo* and *in vitro*. Ideally, they are proteins without direct relatives within humans, which serves to minimise the potential for off target effects. This method of 'rational' drug design, had some notable early successes, leading to the development of sulphonamides and quinolones amongst others. The last new class of antibiotic

discovered *via* this route was the Oxazolidinones in 1987, during a period when > 20 pharmaceutical companies were directly engaged in antibiotic discovery.[24] In addition to developing new antibiotics from first principles, there have also been some successes reported repurposing and optimising known molecules to enable activity against resistant strains of clinical significance.[25–27]

Of all the antibiotics discovered to date, the majority have been isolated from fungi or bacteria. Around half of all commercially licensed antibiotics in use today are based on molecules isolated from *Streptomyces spp.*[24] This relative abundance is due to the wide variety of antimicrobial compounds produced by many bacterial species, which form an important component of the producing microorganism's secondary metabolome. Unlike the products of primary metabolic pathways, these compounds are chemically diverse and are an unequalled source of bioactive potential. For these reasons, such natural products remain one of the preeminent sources of new bioactive drug leads, yielding many molecules with potent antibiotic activities.

## 1.2 Natural Products

Natural products (NPs) include any product or compound produced by a living organism.[28,29] They include both primary metabolites and secondary metabolites, though are generally considered to be comprised predominantly of the latter. Secondary metabolites have diverse roles within plants, animals, fungi and microbes. They are considered to convey a competitive advantage to the natural producer, and hence provide an enhancement in fitness. Secondary metabolites include, but are not limited to, toxins, molecules with roles in intercellular signalling and quorum sensing, defensive compounds, and those necessary for symbiosis and other ecological interactions such as harvesting minerals or metals from the environment.[30] The vast majority of secondary metabolite NPs are predicted to be currently unknown to science, and although the majority of those that have been isolated and characterised are not actively used by mankind, a significant number have been exploited for use in a variety of industries. They are used in the production of fragrances,[31] neutraceuticals,[32,33] in personal care products[34] and as agrochemicals.[35] Importantly, these compounds also represent an abundant source of therapeutic agents, including, but not limited to, antibiotics. Examples include anti-cancer[36–40] and anti-viral[41–44] agents, immunomodulators[45–47] and enzyme inhibitors.[48,49] NPs have also been employed extensively in food preservation.[50] Whilst initial attempts to isolate NPs from their host species resulted in the 'golden age' for antibiotic discovery, the pace at which new compounds were characterised declined rapidly from the 1980s onwards. This was due to the misguided belief that microbes had been exhausted as a source of novel therapeutic lead compounds, a move that coincided with the emergence of the target driven approach for drug discovery.

### 1.2.1 Natural Product Discovery

The last two decades have, however, seen a resurgence in the field of NP discovery and exploitation; a consequence of major improvements in the ease and accessibility of DNA sequencing and allied technologies.[51] These technological advances enabled the sequencing and subsequent construction of the *Streptomyces coelicolor* and *Salinispora tropica* genomes, which revealed that > 70 % of their complement of NP producing biosynthetic pathways were cryptic and previously unknown.[52,53] This observation, in conjunction with a recognition that there are many bacterial species unculturable under conventional laboratory conditions, has led

to recent estimates of the potential existence of > 10 billion unique NPs.[54] The ability to combine genome sequencing with novel approaches for bacterial cell culture[55] and the exploration of ever more remote geographies, with their highly specialised and unique life forms,[56] has led to an explosion in the number and diversity of natural compounds characterised. For example, in 2017 >1490 compounds were isolated from marine microorganisms alone, an increase of >17 % from 2016.[57] These numbers are expected to grow yet further in the coming years, with studies in the Antarctic revealing large numbers of NP encoding genes endemic to the region.[58]

The emergence of AMR, in combination with recent high profile success in the identification of 'new' bioactive NPs[36] has led to a surge in interest in the development of new NP based antimicrobials. Using an innovative microbial culturing device termed the Isolation Chip (iChip),[59] developed in 2010 to enable researchers to "culture unculturable bacteria", Ling *et al*. were able to isolate the NP antibiotic teixobactin. This compound inhibits cell wall biosynthesis,[60] and possesses highly potent bactericidal activity against Gram-positive bacterial strains resistant to vancomycin and methicillin, whilst exhibiting no toxicity against mammalian cells and without the detectable emergence of resistance. Despite repeated passages of *S. aureus* bacteria in concentrations of teixobactin below the minimum inhibitory concentration (MIC) over 27 days, no resistant colonies could be recovered. Culp *et al*. recently used a bioinformatic approach, with the knowledge that genetic information encoding NP-producing enzymes is normally co-located on the host genome, to identify biosynthetic gene clusters (BGCs) without the presence of observable known resistance determinants.[61] This led to the discovery of a novel glycopeptide antibiotic, corbomycin, which also has activity against methicillin, daptomycin and vancomycin resistant strains of *S. aureus* and *Enterococcus*. Corbomycin elicits its bioactivity by interfering with peptidoglycan biosynthesis, but not by blocking an enzyme involved in constructing the cell membrane, rather one responsible for peptidoglycan remodelling.[61] Most recently, Stokes *et al*. reported the discovery of halicin, named after the Heuristically Programmed ALgorithmic Computer (HAL) from the 1968 film 2001 – A Space Odyssey.[62] Halicin is so named due to its discovery using a deep neural network approach. Artificial intelligence was trained in a supervised manner to identify the key attributes for a successful antibiotic. This program then analysed libraries containing over 107 million different chemical compounds. Halicin displays activity against both Gram-positive and Gram-negative bacteria, including against *Acinetobacter baumannii*, which is regarded by the WHO as one of the "priority pathogens"; bacteria for which antibiotics are most urgently required. Halicin has been demonstrated to kill *pan*-resistant *A. baumannii* in mice models,[62] leading to hopes that off-

target effects are minimal. Whilst it is not derived from a natural product, it does highlight the value of computational approaches in modern antibiotic discovery processes. Together, these studies constitute high profile exemplars of the novel approaches being applied to discover novel antibiotics, in particular those of NP origin. These innovative methods have yielded structurally unique antibiotics with previously unreported modes of action and broad spectrum activity against multiple MDR clinical isolates. A lack of discernible resistance when investigated is also extremely encouraging.

Another attractive approach for NP discovery is to encourage the production of NPs from gene clusters that are inactive, or 'silent', under standard laboratory conditions. With the genomes of many well-studied species of *Streptomyces* encoding up to 40 individual gene clusters, each with the potential to produce a distinct secondary metabolite,[54] it is easy to conceive of the vast untapped NP potential of these bacteria. In order to unambiguously establish what compounds a BGC might produce, predict their structure and efficacy computationally, and to produce more chemically distinct derivatives, the enzymology of the biosynthetic pathways used to produce these compounds must be understood. The pathways used to produce NPs are incredibly diverse, due in part to the fact that secondary metabolite production is generally strain-specific.[63] This contrasts with primary metabolism where there is significant conservation within the molecular machineries employed, due to the essential nature of these processes. This diversity means that many BGCs remain unclassified or poorly understood, with many new enzymes and their resultant chemistries yet to be identified.[64]

### 1.2.2 Natural Product Families

A wide range of secondary metabolite families have been identified in bacteria, classified based on commonalities in their chemical origins and biosynthetic pathways used to produce them. The collinearity between nucleic acid sequence and product chemistry often enables prediction of NP structure based on BGC sequence. Saccharides are the most common secondary metabolite produced by bacteria (Figure 1.2), encoded within 93% of bacterial genomes.[64] This is unsurprising due to the universal role that these metabolites play in biology, acting as signalling molecules and forming major components of cell membranes.[65] Terpenes are the most structurally and chemically diverse group of NPs recorded to date; they are produced by both prokaryotes and eukaryotes for a variety of functions.[66] They include carotenoids for photosynthesis[67] and, in humans, bacteria in the gut microbiome are crucial for the production of

retinoids such as Vitamin A.[68] Terpenes are particularly abundant in plants and include the antimalarial compound artemisinin,[69] and the multitude of fragrance compounds found in cannabis resin.[70]

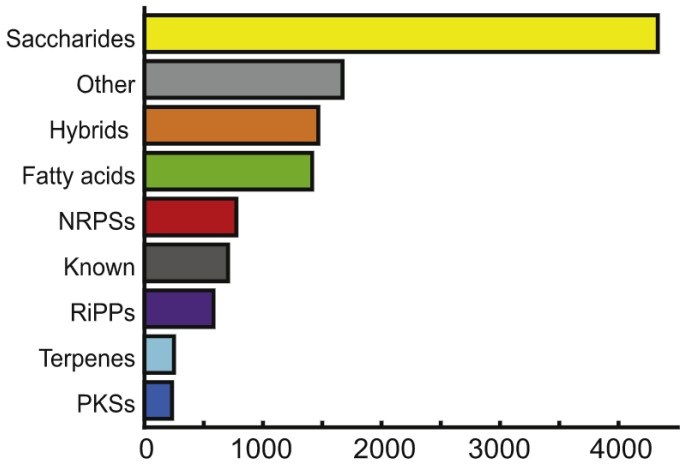



**Figure 1.2 The Distribution of NP Secondary Metabolites in Bacteria.** Over 1150 bacterial genomes were analysed and BGCs identified and categorised by product. 'Other' includes over 1000 BGCs that were unable to be classified, as well as a number of rarer, known secondary metabolite BGCs. In a third of the bacterial genomes analysed, Saccharides made up over half of the BGCs present, whilst they represent over 40% of all BGCs identified. Adapted from Cimermancic et al.[64]

Ribosomally synthesised and *post*-translationally modified peptides (RiPPs) are a less well studied family of NPs, though these molecules possess potential for use across a range of application areas. In these systems, a precursor peptide, attached to a short leader sequence, is biosynthesised by the ribosome. This molecule is subsequently recognised by downstream enzymes which *post*-translationally modify the peptide core to generate the mature metabolite. Whilst the antibiotic thiostrepton is used in complex ointments, the use of RiPPs as therapeutics has been hampered by their poor bioavailability and solubility. This notwithstanding, the promiscuity displayed by RiPP tailoring enzymes shows tremendous potential for use in bioengineering approaches.[71] These enzymes recognise the leader sequence and then act upon the highly variable core of the peptide, creating a structurally diverse suite of biomolecules without the need to specifically recognise the peptide they are acting upon.[72]

With respect to clinically proven therapeutic NPs, non-ribosomal peptides (NRPs) and polyketides (PKs) are the two most exploited families of secondary metabolites. They are currently used as immunosuppressive drugs, anti-cancer agents, and are some of our most effective and well known antibiotics (Figure 1.3). Clinically relevant NRPs include the anti-cancer drug actinomycin D,[73] the antibiotic vancomycin[74] and the antiparasitic avermectins.[75] PKs include the antibiotic erythromycin,[76] whilst hybrid NRP-PKs, which are comprised of building blocks originating from both pathways, have also proven to be exploitable therapeutic agents. Examples include bleomycin, which is used to treat a wide range of carcinomas and lymphomas,[40] and rapamycin and its analogues, which exhibit immunosuppressive, anti-cancer, anti-fungal, anti-aging and neuroprotective activities[47,77]. The NRP-PK epothilone has also found clinical use as an anti-tumour agent, due to its capacity to stabilise microtubules and induce apoptosis.[78] Polyketides and hybrid NRP-PKs systems in particular have proved to be highly successful sources for pharmaceutical drugs.[79] As of 2005 > 20 of these agents were available commercially for clinical use, equating > 0.3 % of all known molecules of this class. This compares extremely favourably to the < 0.001 % success rate reported from target based drug discovery screening programmes.[80] Thus, the case for PKs, NRPs and their hybrids as priority drug leads is a compelling one. In addition to traditional NP discovery methods, there is further hope that synthetic biology approaches will enable the identification and functional optimisation of these compounds.[63] With this as a goal, the field of metabolic engineering aims to construct unnatural biosynthetic pathways, both within cells and *in vitro*, to produce molecules with specific chemistry and function. These approaches hope to counter some of the limitations of synthetic chemistry; using enzymes to perform one pot biotransformations that might take a synthetic chemist many steps to achieve, often with poor yield.[81,82]

## 1.3 Natural Product Biosynthesis

### 1.3.1 Exploiting Natural Product Biosynthetic Pathways

NP biosynthetic pathways are considered attractive targets for exploitation and re-engineering, often outside of their natural context. For example, RiPP tailoring enzymes have recently been shown to be highly promiscuous, leading to the creation of non-natural hybrid RiPP products.[83] This has enabled the assembly of chimeric peptides with leader sequences specific for a plethora of tailoring enzymes, including those not encoded within the natural pathway. The tailoring enzyme then recognises its cognate leader sequence and acts upon its fused partner peptide. This capacity for a mix-and-match approach shows promise accessing rationally designed peptide products with predefined functionality. Similarly, non-ribosomal peptide synthetases (NRPSs) and polyketide synthases (PKSs) are both considered attractive targets for re-engineering, due in part to their distinctive modular architectures. Both use predictable and conserved classes of enzymes to select and install starter and extender unit substrates, assembling a product chain *via* sequential, *pre*-programmed condensation reactions, in a manner analogous to fatty acid biosynthesis.[84] The resulting product backbone is then liberated from the assembly line-like enzyme and further modified by downstream tailoring enzymes. The level of conservation within these classes means that both NRPS and PKS gene clusters can be readily identified *in silico*. Genome sequencing data can then in turn be used as a proxy to predict the chemistry of the final pathway product, by aligning the sequence with the sequence of known BGCs.[85] This approach is, however, not without its limitations. Key pharmacological nuances, such as compound stereochemistry, or compound modification by *trans*-acting enzymes are sometimes impossible to predict.[86]

PKSs and NRPSs differ principally in the substrate starter and extender units that they use (Figure 1.4). PKSs incorporate predominantly malonyl- and methylmalonyl-CoA building blocks, whereas NRPS enzymes have specificity for a much wider range of amino acid substrates. These include all 20 naturally occurring amino acids, in addition to many hundreds of non-proteinogenic and unnatural amino acids, and aryl acids.[87,88] PKSs, which are the focus of the work outlined in this thesis, may be classified into one of three distinct families, termed type I, II, and III, which differ in their molecular architectures and modes of action (Figure 1.5).

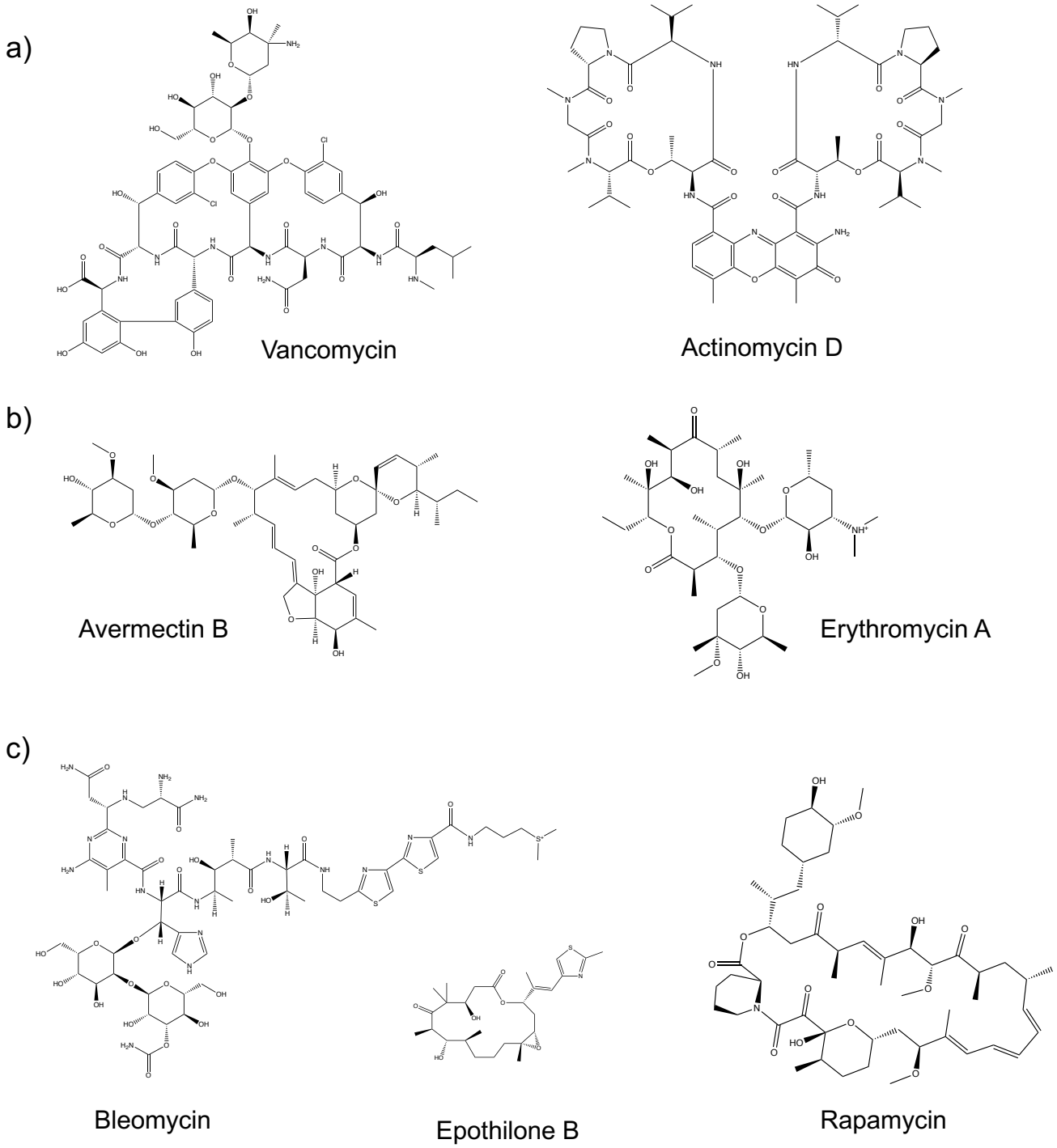



**Figure 1.3 Biosynthetically Constructed, Therapeutically Relevant, NPs.** a) NRPS derived therapeutic NPs. b) PKS NPs. c) Molecules produced by hybrid PKS-NRPS enzymatic assembly lines.

### 1.3.2 Type I Polyketide Synthases

Type I PKSs are the largest and most sophisticated family of NP biosynthetic enzymes and are broadly divided into two subgroups, iterative PKSs (iPKSs) common to fungi, and modular PKSs (mPKS), found predominantly in bacteria. A detailed discussion of Iterative systems goes well beyond the scope of this thesis, and thus these systems will not be discussed here.

mPKS are encoded by large open reading frames (ORFs), that encode for gigantic multi-domain polypeptides. In most instances each domain within the synthase acts only once during the biosynthetic possess, rendering these systems inherently non-iterative. Individual PKS domains are arranged into distinct modules within the PKS, with each module minimally composed of three domains, a ketosynthase (KS), an acyltransferase (AT), and an acyl carrier protein (ACP). The AT is usually highly selective, functioning as a gatekeeper by dictating the choice of acyl-CoA derived substrate to be loaded at each module. Catalytically, these enzymes employ a nucleophilic serine residue, housed within the enzyme active site, to attack the thioester of the selected acyl-CoA substrate, facilitating substrate transfer to the neighbouring ACP via a ping-pong mechanism.[89] KS domains are responsible for catalysing Claisen-like condensation reactions, which link the AT-selected extender unit to the growing polyketide chain. The ACPs act as way stations, presenting either the AT-loaded extender units or the extended product chain for downstream modification. Both substrates and intermediates are attached to ACPs *via* a phosphopantetheine arm, *post*-translationally added to a conserved serine residue, which attaches the acyl-O-enzyme intermediate via its terminal thiol group.[88,90] Each module thus serves to extend the product chain by two carbon atoms. Following extender unit incorporation, the β-branched ketone of this substituent can be selectively reduced to an alcohol, alkene or alkane through the action of module specific ketoreductase (KR) domains, dehydratase (DH) domains, and enoylreductase (ER) domains, respectively. Each module has a defined compliment of reductive domains, which allows defined control over the oxidation state of product incorporated monomers. In addition to these core domains, terminal modules within PKSs often incorporate a thioesterase (TE) domain, which acts to liberate the immature polyketide product from the synthase assembly line.

The paradigm example of this type of canonical mPKS is the 6-deoxyerythronolide B synthase (DEBS).[91] It has been studied extensively for over 50 years and is described in greater detail below. Although a multitude of synthases have been recently reported that deviate from this prototypical canonical architecture, most notably the *trans*-AT family of PKSs,[89] these are

of only tangential relevance to the work outlined herein and will thus not be described in significant detail in this introduction. Importantly Polyketide synthases of the *cis*-AT class, which adhere to the formal principles outlined above are considered to conform to the 'Principal of Collinearity'. This dictates that the chemistry of the synthase product itself is defined by the order and composition of the PKS assembly line, which is in turn dictated by the nucleotide sequence which encodes the biosynthetic machinery of the pathway.[92]

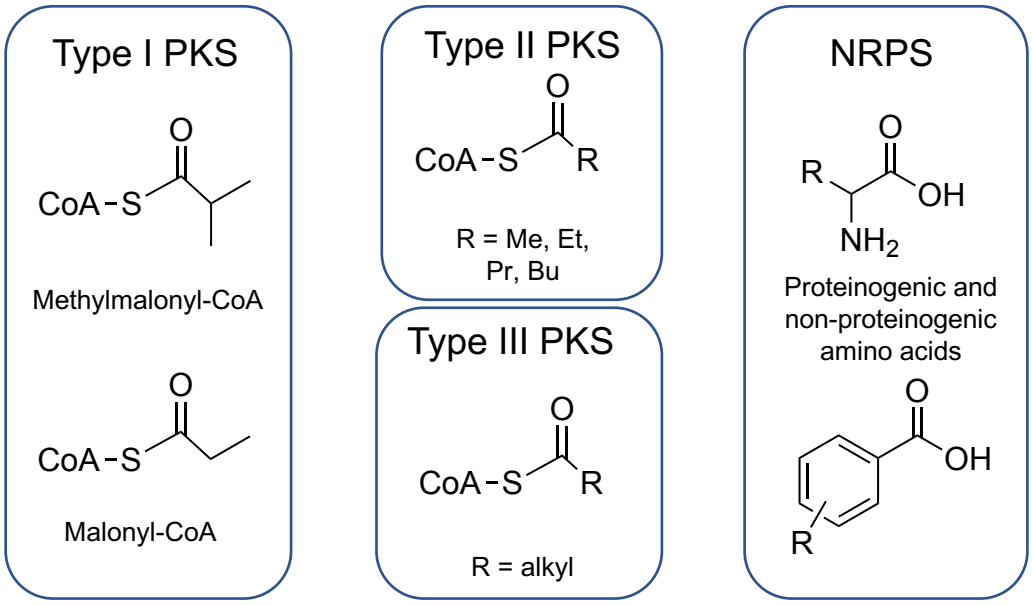



**Figure 1.4 Commonly Used Starter and Extender Molecules in Polyketide and Non-Ribosomal Peptide Synthesis.**

In contrast to Type I mPKSs, Type I iterative systems act in a cyclical manner, assembling their polyketide product using a single set of enzymatic domains to perform multiple rounds of chain elongation, typically using malonyl-CoA extender units (Figure 1.5). These systems are similarly able to control the degree of product reduction through the action of KR, DH and ER domains, but in contrast to mPKS enzymes, these modifications take place during predefined cycles of chain extension.[93] Importantly, studies have reported the presence of semi-iterative and iterative modules embedded within mPKSs, in addition to the presence of non-canonical domains, casting doubt on the accuracy of making predictions of product chemistry based purely on synthase nucleotide sequence.[94]

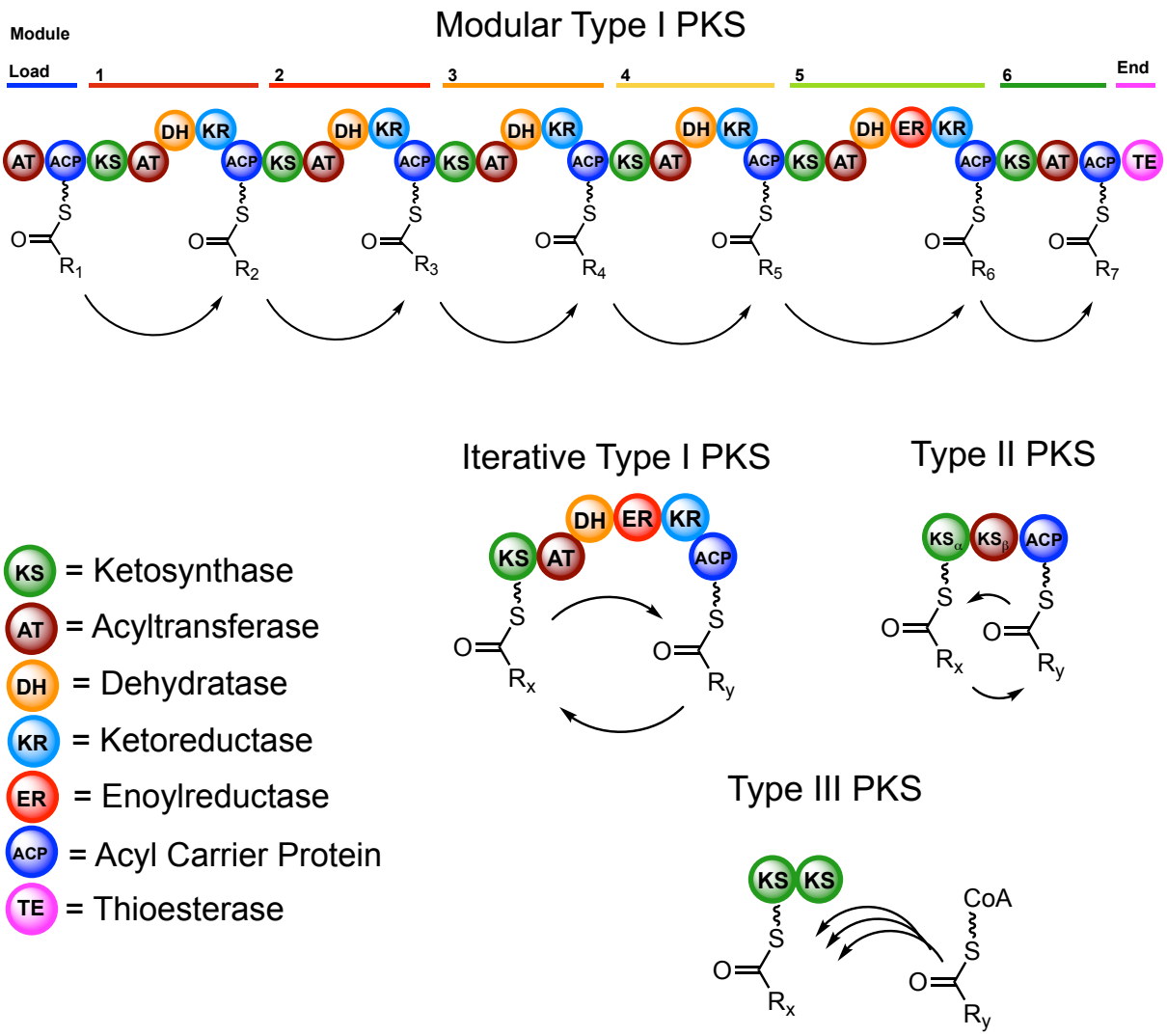



**Figure 1.5 Schematic Representation of Polyketide Biosynthesis.** Each domain is represented by a labelled circle. Type I mPKSs consist of many modules split amongst long polypeptide chains, which could each contain 1 to 4 modules. Type I iPKSs consist of a single 6 domain module. Type II PKSs also work iteratively, although the domains are expressed individually. Type III PKSs do not interact with ACPs in order to create the polyketide chain, instead interacting directly with acyl-CoA.[95]

### 1.3.3 Type II Polyketide Synthases

Type II polyketide synthases are responsible for the biosynthesis of an array of bioactive aromatic polyketides, including the anticancer anthracyclines. They are almost exclusively found in actinomycetes bacteria.[96] The principal difference between type II and type I systems

is that type II systems comprise a series of dissociable enzyme domains, which are the products of individual ORFs. A minimal type II module consists of an ACP, which anchors the growing polyketide chain, and two KS subunits, $KS_\alpha$ and $KS_\beta$ (also known as the chain length factor, CLF). The two KS subunits strongly associate to form a heterodimer, and link to the ACP, with the $KS_\alpha$ domain catalysing the Claisen condensation reaction necessary to extend the product chain.[97] The $KS_\beta$ domain is the primary determinant of the length of the product chain synthesised and forms a cavity with the $KS_\alpha$ domain which is, in turn, occupied by the nascent polyketide chain. This amphipathic tunnel simultaneously prevents any spontaneous cyclisation and strictly controls the final chain length.[98,99] Although the majority of type II PKSs exclusively use malonyl-CoA as the starter and extender units, there have recently been reports of systems which are capable of using alternative starter units, such as 5-carbon 2-methylbutyryl-CoA in the biosynthesis of trioxycarcin.[100] Once assembled, Type II PKS products are then further modified by downstream enzymes including aromatases, cyclases, oxygenases and transferases of various flavours, to decorate the aromatic polyketide core.[101]

### 1.3.4 Type III Polyketide Synthases

Polyketides produced by Type III systems are typically found in plants, although it is now accepted that there are also widespread examples to be found in bacteria.[102] They are structurally the smallest and simplest of the PKSs, conducting all reactions necessary to construct the polyketide product in a single active site. This exists within each monomer comprising the functional homodimer.[103] Similarly to other PKSs, they iteratively catalyse Claisen condensation reactions, building a polyketide chain from acyl-CoA building blocks, usually malonyl-CoA. Intramolecular cyclisation catalyses the release of the completed polyketide chain from the PKS. Despite their simplicity, type III PKS products are highly diverse and exhibit a wide range of bioactivities.[95] This includes their integral role in mycobacterium biofilm formation and pathogenesis.[104] The structural and functional diversity of these molecules originates in the variety of starter and extender units employed during biosynthesis and the nuances of the chain cyclisation process, with this latter regulated by the shape of the active site cavity.[95] Notably, extender units used by the Type III system can be widely varied (Figure 1.4), for example, $C_{22}$-$C_{26}$ FAS products are used as starter units for the production of phenolic lipids in *Azotobacter vinelandii*. Crucially, the FAS established a highly specific protein-protein interaction with the PKS to enable starter unit transfer.[105]

### 1.3.5 Non-Ribosomal Peptide Synthases

NRPSs are analogous to type I mPKSs in that they are composed of large multi-domain polypeptides that operate in a non-iterative fashion to assemble a NP scaffold, although NRPS biosynthesis relies on the use of amino acid, rather than carboxylic acid, building blocks. Minimally, a single NRPS module consists of the adenylation (A) domain, a peptidyl carrier protein (PCP) domain and a condensation domain (C). The A domain functions in a manner analogous to a PKS AT, selecting the extender or starter unit and loading it onto the terminal thiol of a phosphopantetheine arm attached to an ACP equivalent PCP domain. The C domain then catalyses the formation of a peptide bond between the *pre*-formed peptide intermediate and loaded amino acid extender unit, resulting in chain elongation, in a manner akin to a PKS KS domain. The chain can then be further modified by accessory domains including methyltransferases and epimerases, which can function to modify the stereochemical output of the chain extension process.[106]

### 1.3.6 The DEBS Paradigm

The work outlined in this thesis is primarily concerned with bacterial type I mPKSs, the tailoring enzymes that they employ, and the NPs that they produce. In order to establish a foundational understanding of the core principles of mPKS enzymology it is necessary to illustrate these features using a paradigm exemplar. The DEBS polyketide synthase, which biosynthesises the 6-deoxyerythronolide B (6-dEB) core of erythromycin, is most studied and best understood of the mPKSs.[90] In the early 1990s multiple high-impact papers concerning the domain architecture and modular organisation of DEBS were published, along with successful expression studies of DEBS within heterologous hosts, including *E. coli*.[107–110] These findings, and subsequent clarifications, are summarised diagrammatically in Figure 1.6. They represent the first successful attempts to provide a comprehensive understanding of PKS enzymology at a molecular level. Importantly these studies themselves build on decades of prior research exploring the molecular genetics of the natural erythromycin producer *Saccharopolyspora erythrea*.[111,112]

DEBS is constructed from three large polypeptide chains, each encoded by its own ORF; DEBS1, DEBS2 and DEBS3. These three protein chains assemble a single 14-carbon polyketide product, employing a primary loading domain on DEBS1, two extender modules on each

of DEBS1, 2 and 3, and an additional terminal TE domain on DEBS3. Each module further contains a KR domain, although the KR domain of module 3 is catalytically inactive. Module 4 contains additional DH and ER domains, resulting in a fully reduced portion of the polyketide backbone. The TE domain of DEBS3 facilitates the liberation of the immature polyketide chain from the assembly line, *via* cyclisation, to form the 6-dEB macrocycle. In summary, 28 active sites are arranged across 29 enzyme domains within three large polypeptide chains. This foundational understanding of the molecular organisation of DEBS has been used to guide efforts in the determination of molecular structures, interactions and kinetic processes for multiple type I mPKSs. For example, it is now known that KS-AT didomains assemble to form homodimeric protomers, in a manner dictated by the KS domain, with key intramolecular protein-protein interactions observed between domains.[91,113,114] It is also known that linker regions are required to control module–module interactions both within and between individual polypeptides. α-helical docking domains allow the highly specific passage of the polyketide chain between adjacent modules.[115,116] Further flexible linkers differentiate individual domains within polypeptides, facilitating substrate-induced conformational changes. Investigation into the rules defining the substrate specificity of individual AT domains have been undertaken, in addition to studies to establishing the basis of AT-ACP complementarity.[89,117]

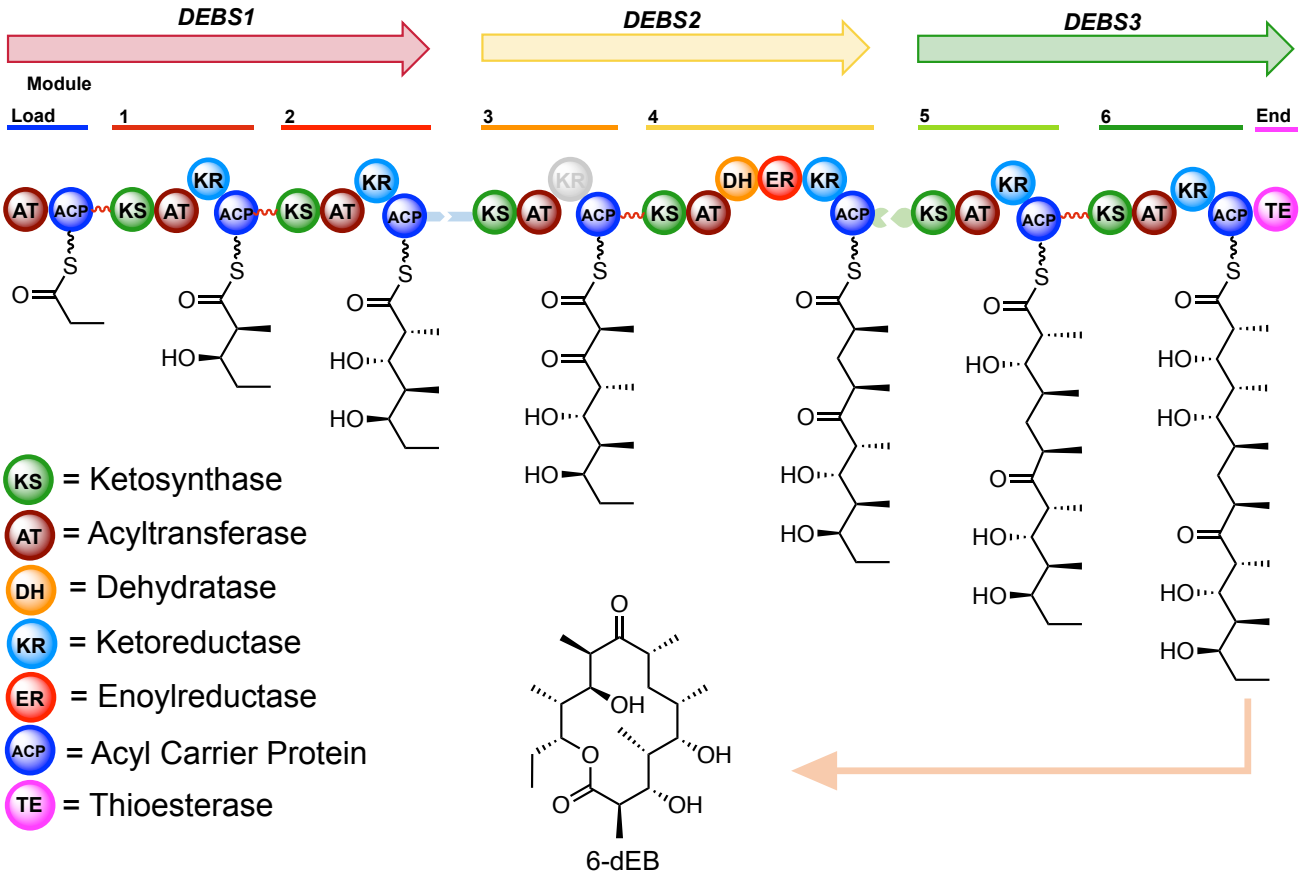



**Figure 1.6 Schematic Diagram of the DEBS mPKS Architecture.** *Intra*-polypeptide linkers are shown in red, whilst the *inter*-polypeptide docking domains are shown in blue and green.

## 1.4 Engineering Natural Product Assembly Lines

### 1.4.1 "Legoization" of Polyketide Synthase Domains

The insights provided by these detailed enzymological studies of DEBS served to drive efforts towards the rational engineering of this synthase, aiming to access a functionally optimised variant of erythromycin by interchanging domains and modules. This approach has been described as the "Legoization" of mPKSs.[118] The principle idea behind this approach is that each domain might be used much like a LEGO® building block, combined with other domains to lead to the construction of a designer PKS, biosynthesising similarly bespoke polyketides. Ultimately, it is hoped that such an approach could lead to the establishment of huge libraries of random combinatorial polyketides. Indeed, theoretical studies show that by only using chemistry native to the synthesis of 6dEB, it should be possible to generate >100,000 possible chemical structures, once the number of elongation steps, possible starter molecules, possible extender molecules and their possible chirality, and the level of reduction is taken into account.[119] Once all naturally occurring chemical possibilities present in other known pathways are taken into account, this scales to > 800 million theoretically possible polyketides, of which only 10,000 had been reported as of 2005. Another potential avenue of exploration lies in the prospect of using a semi-synthetic semi-biosynthetic approach to further diversify these compounds. A chemo-biosynthetic approach should in principle lead to an even greater chemical diversity.

There were some notable early successes re-engineering DEBS, with McDaniel *et al*. producing a library of over 50 different erythromycin analogues in 1999, by using single, double and triple domain alterations. These relied predominantly on AT substitutions, insertions of additional reductive domains and/or removal of reductive domains.[120] The yield of these analogues was, however, noticeably lower than for the authentic natural product, with yields so low that product mass determination proved problematic. Further progress was made by Menzella *et al*., who successfully cloned 14 separate modules from 8 different mPKS clusters and assembled them into > 150 different combinations.[121] *In vivo* analysis in *E. coli* allowed determination of those combinations that were biosynthetically active, resulting in the isolation of a number of novel unnatural natural products. Fewer than half of the constructs produced detectable product, however, and all were at a lower yield than the parent PKSs. Unfortunately, the low yields limited the usefulness of these approaches in commercial drug discovery.[80]

### 1.4.2 Engineering Acyltransferase Domains

Significant efforts were made to swap AT domains between mPKS modules, as this represents an inherently attractive route to accessing novel products. If an alternative extender module can be inserted into a polyketide backbone at a defined position, the prospects for accessing new natural product chemistry increase significantly. Initial efforts sought to exploit the natural promiscuity exhibited by the loading module of the avermectin biosynthetic pathway, which installs both 2-methylbutyryl-CoA and isobutyryl-CoA to generate avermectins A and B. Early feeding studies showed that this AT could incorporate > 40 different starter units,[122] and that some of this promiscuity could be readily translated into the DEBS system.[123] Other studies successfully used point mutations to knock out *cis*-AT activity in DEBS, replacing this essential function with surrogate *trans*-acting AT domains from other systems. This approach has delivered successes both *in vivo*,[124] and *in vitro*,[125] and mostly without appreciable loss of activity. Other researchers have succeeded in using this complementation approach to insert fluorine atoms into erythromycin analogues *in vivo*.[126] These studies served to set the benchmark for PKS engineering, though there have been more failures than successes and the true potential of PKS re-engineering is yet to be fully realised.

In addition to conducting complete AT domain swaps between PKSs, efforts have also focussed on attempts to selectively alter the substrate specificity of these enzymes. The exact mechanism by which ATs achieve substrate selectivity for malonyl- or methylmalonyl-CoA is currently not fully understood, and nor is the basis for AT substrate promiscuity.[127] For example, in the lipomycin PKS system, the initial loading and extender modules exhibit remarkably broad specificity for a number of propionyl analogues, but are unable to accept malonyl-CoA as a starter or extender unit.[128] Such ambiguities may, however, be resolved, through the provision of a detailed molecular description of these AT domains, both alone and in the context of the PKS modules within which they reside.

### 1.4.3 Obstacles to Polyketide Synthase Engineering

Despite some successes, the frequent failure of efforts to achieve the problem-free substrate transfer and assembly of PKs by non-natural PKS assembly lines only serves to highlight the oversimplified initial view of PKS enzymology. In reality individual domains have preferred upstream and downstream partners, and distinct substrate specificities.[80] The mechanism by

which this selectivity is conferred is influenced by a multitude of factors at the domain, module and intact PKS level. It has been shown, for example that successive PKS polypeptides use terminal α–helical docking domains to form highly specific complexes with their partner poylpeptides.[129] Any alteration of these sequences results in dramatically reduced product yields, as do attempts to mismatch these key elements.[115] Chen *et al*. demonstrated that they could improve non-natural module activity by preserving the interactions within KS-AT didomains, using alternative ACP-substrate partners. This resulted in catalytic activity at a level of 70% of that of the wild type module.[128]

In other attempts to assemble individual modules *de novo*, extensive module skipping was observed. This phenomenon occurs when the polyketide chain is transferred directly between ACP domains, bypassing KS and associated reducing domains, and delivering truncated polyketide products.[130,131] It is apparent that a failure to account for the contributions made to the biosynthetic processes by complex conformational interactions is likely to translate into aberrantly functioning engineered PKSs. Further, and in light of these observations, it is also clear that in order to successfully apply a mix and match approach by swapping domains and modules between systems, a much more comprehensive structural understanding of mPKSs is needed.[132]

## 1.5 The Structure of Polyketide Synthases

### 1.5.1 The Architectures of Individual Domains

Initial structural studies of mPKSs relied on the use of limited proteolysis experiments to isolate the constituent parts of the DEBS PKS, which could in turn be subjected to high resolution structural studies.[133] The first mPKS domain to be successfully crystallised was the TE domain from DEBS, with the structure of this protein subsequently determined by single crystal X-ray diffraction.[134] This structure revealed a dimerised domain architecture consistent with the known oligomeric state of the intact mPKSs. The 1.79 Å crystal structure of the KR domain from DEBS module 1 was the next structure to emerge.[135] This was rapidly followed by crystal structures of two KS-AT didomains. The first of these structures was from DEBS module 5,[113] which was followed by the structure of the equivalent didomain from DEBS module 3.[114] These were particularly significant, as they gave insights into both individual domain architectures and interdomain protein-protein interactions. Both structures revealed homodimeric complexes with an extensive dimerisation interface between the two KS monomers. They also disclosed the nature of the KS-AT linker regions and the α–helical docking domain on the N-terminus of module 5 (Figure 1.7a). These structures show that the AT domains are arranged horizontally, relative to the central KS dimer, implying that major structural reorganisation was required to enable the ACP domains to engage both the AT and KS domains.[91] Interestingly, these structures were consistent with equivalent structures reported from porcine FAS,[136] and led to the theory that their modular architectures could be consistent.

The structure of the first mPKS ACP domain, that from DEBS module 2, was determined by nuclear magnetic resonance (NMR) spectroscopy. This approach was used in preference to X-ray crystallography due to the modest size (~10 kDa) and flexibility of these domains.[137] To date, no crystal structure incorporating a free ACP has been determined, although the structure of an ACP in complex with its cognate AT, enabled through the use of chemical cross linkers, has been reported.[138] The structure of the DH domain of DEBS module 4 was reported in 2008, again showing notable similarities to mammalian FAS.[139] A structure of the final core mPKS domain was reported in 2012 by Keatinge-Clay and co-workers, who reported an ER-DH domain crystal structure from module 2 of the spinosyn biosynthetic pathway.[140] This time a notable divergence from the structure of the FAS ER was observed, with a deviant dimerisation orientation between ER monomers revealed. These investigations allowed the construction of

a widely accepted model for the organisation of a complete mPKS module (Figure 7a). Importantly, however, this was derived from structures of isolated domains and not informed by detailed studies of mPKS intact modules.

### 1.5.2 The Architecture of Intact Polyketide Synthase Modules

In 2014 the first structures of intact modules from mPKS systems were published. The first of these studies employed cryo-electron microscopy (cryo-EM) to elucidate the architecture of PikAIII from the pikromycin synthase.[141,142] The second study used solution X-ray scattering (SAXS) to elucidate the low resolution structure of DEBS module 3 fused to the DEBS TE domain, and that of DEBS modules 5 and 6, similarly fused to its cognate TE domain.[143] The DEBS SAXS structures, elucidated by Khosla and co-workers, were consistent with their previously reported crystal structures. The researchers used a rigid body fitting approach to formulate a pseudoatomic model of both the DEBS complexes investigated (Figure 5b). The SAXS structure was in sharp contrast to the Cryo-EM structure of PikAIII. This structure was determined to a maximum resolution of 7.3 Å, insufficient to derive a native atomic model, but appropriate for close fitting of DEBS-based homology models of individual PikAIII domains.[141] This structure revealed an arch-shaped symmetric conformation, with the KS dimer located at the apex of the complex and the AT and KR domains extending vertically downwards. The two ACP domains were found to occupy one of two distinct locations a solvent exposed cavity within the core of the complex. These conformations occurred in an approximate 1:1 ratio, engaging either the KR or AT domains (Figure 5c). In a companion paper, Skiniotis, Smith, Sherman and co-workers also interrogated the structural rearrangements that accompany the different stages of polyketide chain elongation during the PikAIII's catalytic cycle.[142] This was achieved by incubating recombinant PikAIII with its pentaketide substrate, both alone and in conjunction with methyl-malonyl-CoA and/or NADPH. In this way, the researchers were able to elucidate a number of PikAIII structures in a series of intermediate catalytic states. They solved structures with substrate bound to the KS and bound to the ACP both *pre*- and *post*-reduction by the KR domain. Multiple conformational changes were observed, including flipping of the KR domain and movement of the ACP within the central cavity of the module. Subsequent SAXS studies from the Weissman lab, focusing on intact modules from *trans*-AT mPKSs have since established that these PKSs also adopt arch-shaped conformations

in solution,[144,145] although it remains unclear as to how comparable these studies are given the established differences between *cis*-and *trans*-AT systems.

### 1.5.3 Interpreting Structure-Function Relationships in Polyketide Synthases

The Cryo-EM studies of Whicher *et al*. provide compelling support for the hypothesis that protein-protein interactions and conformational reorganisation are essential components of mPKS module catalysis, consistent with the turnstile mechanism previously proposed by Khosla, Cane and others.[146] This theory offers an explanation as to why, during the vectoral synthesis of polyketides, there is no stuttering or skipping as is observed in engineered systems. It proposes that the linkers and interfaces between individual domains act to instigate conformational changes as opposed to playing a purely structural role.[79] In this way, conformational changes prevent the loading of the KS with a new polyketide prior to completion of an in progress catalytic cycle.

Together, these studies serve to pose more questions regarding mPKS module function than they provide answers. Can the PikAIII structure be considered a universal model for mPKS architecture given its apparent disparity with SAXS studies of other mPKS polypeptides? Is the horseshoe topology observed in the PikAIII structure not seen in KS-AT didomain X-ray structures as these latter proteins lack necessary ACP directed protein-protein interactions to adopt this conformation? Would the SAXS data, if refined using the PikAIII cryo-EM structure as a starting model, give a structural description more closely mirroring this architecture?[147] A number of excellent reviews have debated relative merits and apparent disparities between each of these studies, with each arriving at the same conclusion: that further structural information is urgently required to definitively confirm the structure and mode of action of mPKSs.[79,82,147–151]

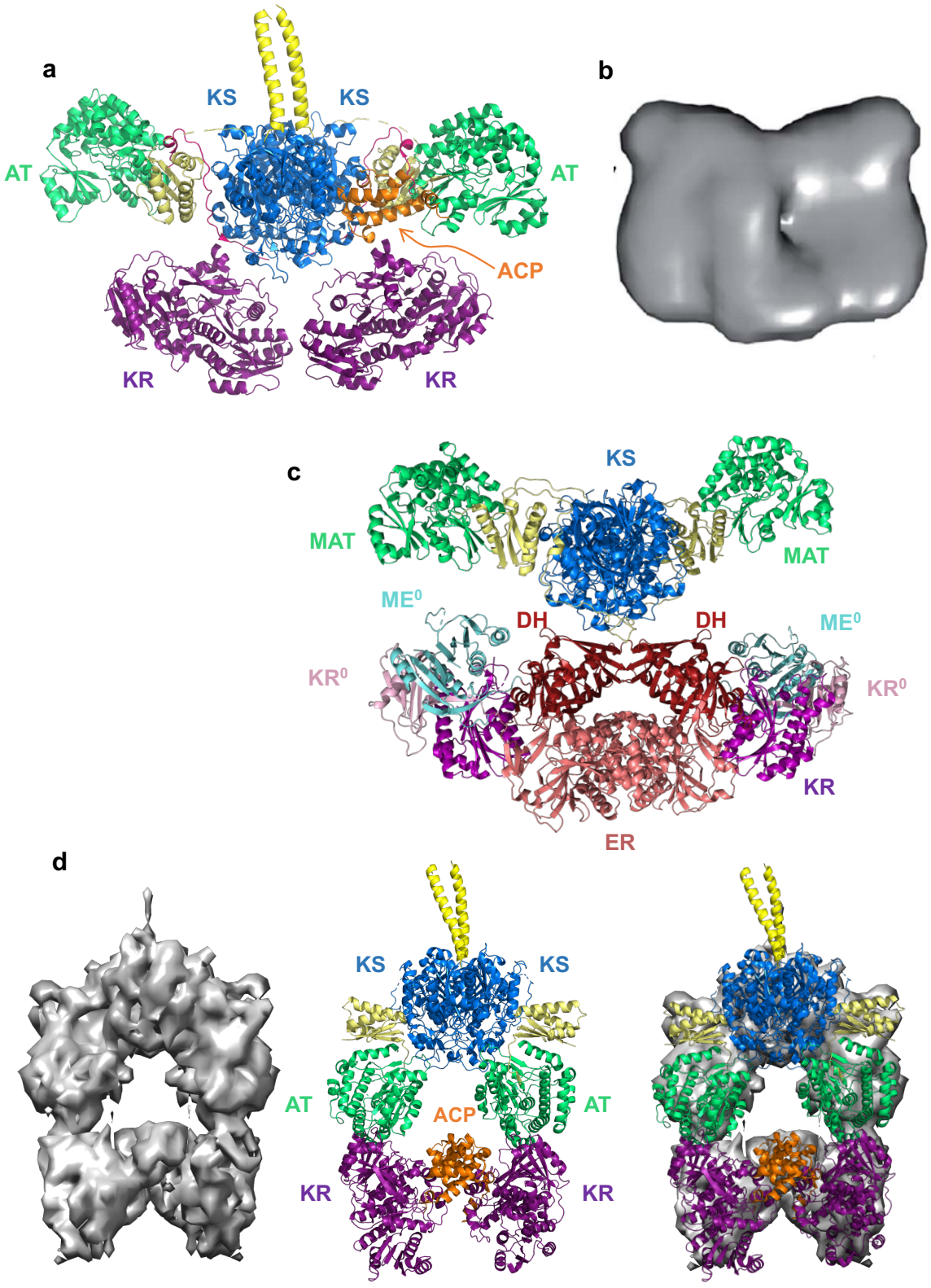



**Figure 1.7 Alternative mPKS Structural Models.** a) 2009 hybrid DEBS modular structure based on structural analysis by the Khosla and Keatinge-Clay groups (based on Khosla et al.[132]). The KS-AT dimer of DEBS module 5, solved by X-ray crystallography, including pre-KS docking domain

(bright yellow), the flexible KS-AT linker (light yellow), and the post-AT linker (pink) forms the core of the assembly. The NMR resolved structure from DEBS module 2 and the KR domain from module 1 were docked to this computationally. Only one ACP domain is shown for clarity. b) SAXS structure of DEBS module 3 connected to a TE domain.[143] c) Structure of the porcine FAS solved by Maier et al. to a 3.2 Å resolution.[152] d) Cryo-EM structure of PikAIII and the molecular DEBS-based atomic map fit to the structure.[141] Colouring is consistent to the Khosla model, with the pre-KS docking domain in bright yellow and the KS-AT linker in light yellow. The post-AT linker is not resolvable and hence is not shown. $KR^0$ = pseudo-ketoreductase (facilitates dimerisation but is inactive), MAT = malonyl-acetyl transferase (analogous to AT in PKS), $ME^0$ = pseudo-methyltransferase (inactive and homologous to methyltransferases).

### 1.5.4 Biochemical Characterisation of mPKS Domains

In tandem with structural studies of mPKSs, biochemical and biophysical investigations have also served to further our understanding of mPKS enzymology. These have included determining the stereoselectivity of a variety of isolated mPKS domains. For example, Valenzano *et al*. investigated the DH domain of DEBS module 4, demonstrating that it is unable to process substrates with alternative stereochemistries, such as those generated by KR domains within non-cognate DEBS modules or from alternative mPKSs.[153] Likewise, ER domains from a number of pathways have been found to exhibit absolute stereoselectivity for their DH generated substrates; though exceptions do exist, such as ER4 from DEBS.[154] Similarly, *in vitro* studies of ACP and AT domains have established some of the determinants of substrate and partner domain specificity, and established kinetic parameters for substrate selection and transfer.[89]

## 1.6 Recent Advances in mPKS Understanding

### 1.6.1 Protein-Protein Interaction-Based mPKS Engineering Approaches

In light of recent progress in the structural and functional characterisation of mPKSs, re-engineering efforts have refocused to account for these important observations. A number of groups have opted to attempt to engineer PKS pathways without overtly interfering with critical protein-protein interactions. This typically involves active site mutagenesis, and results in small, but positive outcomes. Sherman and co-workers successfully used this approach to modify the TE domain from PikAIV to catalyse the macrocyclisation of an unnatural, but closely related, substrate, when attached to PikAIII.[155] Similarly, the specificity of a KS domain from the mycolactone biosynthetic pathway was relaxed using structure guided mutagenesis,[156] while Jianting Zheng's lab has successfully probed the stereoselectivity of ER and AT domains using similar approaches.[157,158] This approach is not, however, a guarantee of success, as is illustrated by the disabling of KR, DH and ER domains in the premonesin biosynthetic pathway.[159,160] Whilst a library of 22 non-natural metabolites was produced, the highest titre was ~10 times less than for the wild type synthase.

AT exchanges have also been explored in greater detail recently, with Yuzawa *et al*. successfully delineating the precise boundaries of the AT domain *in vitro*, allowing for the most efficient translocation of an AT domain to date.[161] This method attempts to relocate the precise AT linker region, allowing for function of the AT without impacting KS activity. This culminated in a series of AT domain swaps that didn't suffer from compromised productivity. The Williams lab has reported the successful creation of a chimeric mPKS which sequentially installs two non-natural extender units.[162] This study was significant, as previous attempts at to achieve sequential modifications only generated systems producing stalled intermediates, which could not be processed by down-stream modules within the PKS. This study exploited the domain boundary prediction approach of Yuzawa *et al*., but despite this still delivered a system that showed compromised product yields vs. the non-engineered synthase. These methods were applied to enable AT domain swaps within PikAIII and PikAIV, investigating the inherent substrate promiscuity of the domains within these modules.[162] Similar substitutions could also be made into DEBS module 6, which is now known to be highly substrate promiscuous as compared to other modules. These experiments revealed the important roles played by the KS and KR domains in substrate selection.

These studies are also driving investigations into the potential for producing commodity chemicals and biofuels using engineered mPKSs.[127] Hagan *et al*. have successfully combined domains from 6 separate mPKS modules to produce adipic acid.[163] Notably, however, attempts at creating true chimeras *in vitro* have still been generally beset by reduced turnover rates as compared to wild type synthases, implying that issues associated with the disruption of protein-protein interactions are yet to be universally mitigated.[166]

### 1.6.2 High-Throughput Polyketide Synthase Re-Engineering

In addition to structure-guided approaches, there have also been a number of attempts at randomised engineering of NP assembly lines. Whilst these high-throughput methods are logistically challenging to implement, there have been some notable successes, albeit mostly confined to NRPSs.[164,165] For mPKSs, Kasey *et al*. have developed a transcription-factor based reporting system to detect the expression of macrolides, such as erythromycin.[166] Wlodek *et al*. have used a recombination-based method to generate libraries of rapamycin analogues *in vivo*, some of which display modified biological activities.[167]

### 1.6.3 Reclassifying Polyketide Synthase Modules

In tandem with these re-engineering studies, researchers have also begun to re-evaluate the fundamental classification of PKS modules and the domains therein. Keatinge-Clay, building on the work of Zhang *et al*. exploring the evolutionary origins of mPKSs,[168] has become a proponent of an alternative modular description of mPKSs (Figure 1.8). This is derived from studies mapping the evolutionary movement of PKS modules between pathways, with the AT typically moving with the downstream KS as opposed to its upstream neighbour.[169] Initial indications are that the adoption of this description is translating into more efficient module swapping, enabling the creation of multiple active venemycin-pikromycin chimeras.[170] It remains to be seen, however, whether this alternative modular arrangement will expedite further engineering efforts and overcome existing bottlenecks in the *de novo* design of mPKSs.

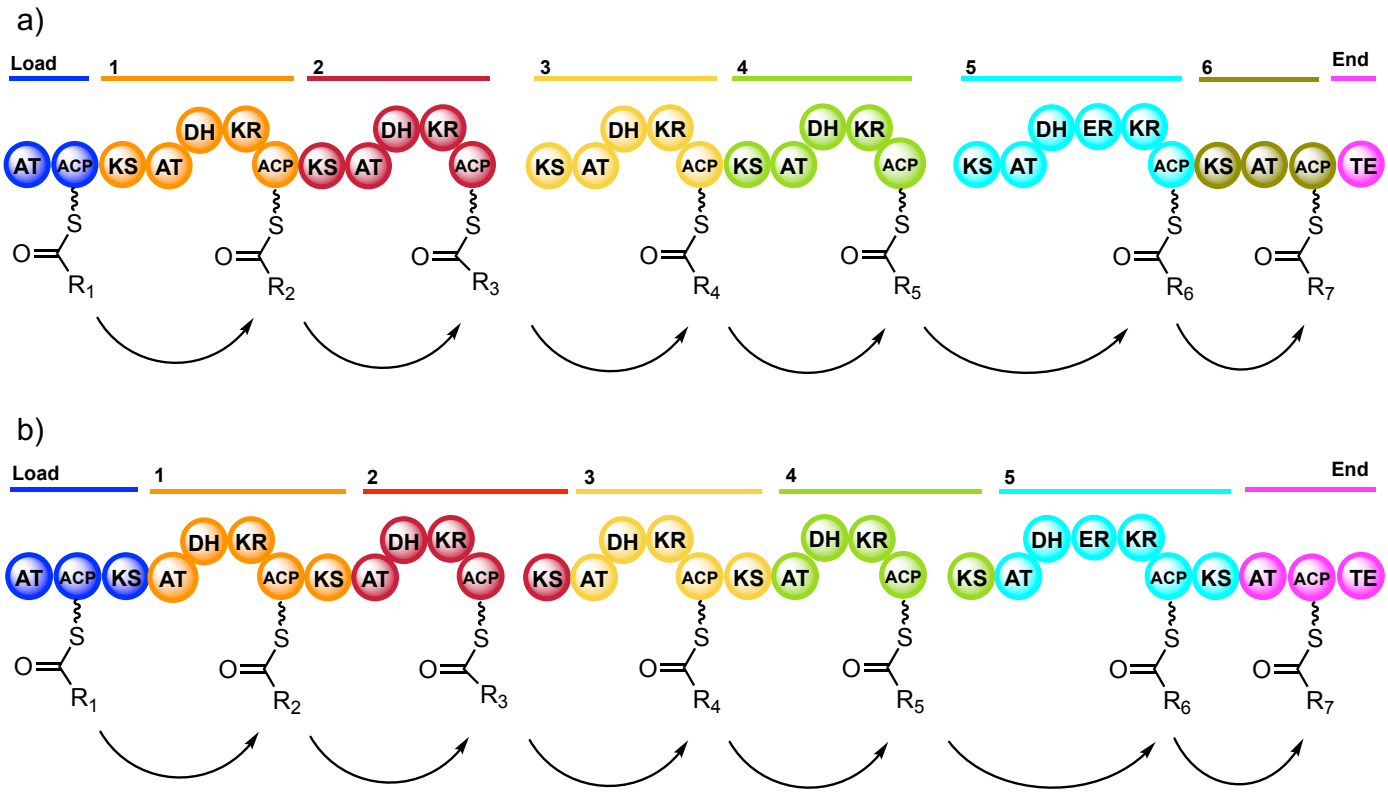



**Figure 1.8 Alternative mPKS Module Classification Schemes.** Schematic representations of an mPKS assembly line, with every domain in a module coloured identically. Distinct polypeptides are represented by gaps. a) Traditional module assignment, in line with polypeptide chains. b) Module assignment proposed by Keatinge-Clay. Modules are spread between polypeptides with the KS being evolutionarily attached to the AT domain preceding it.

### 1.6.4 Future Goals for Polyketide Synthase Research

Efforts to rationally design and build mPKSs appear destined to continue at pace over the coming years. Our capacity to build mPKSs from first principles is still some way from being a tractable proposition with a number of fundamental questions regarding the enzymology of these systems still to be satisfactorily answered. A unified structural description of an intact module is still needed, with the acquisition of additional structural information viewed as a high priority. Ideally such studies should employ a hybrid approach leveraging complimentary methods, including cryo-EM, crystallography, biological NMR and solution scattering. Beyond this, the structure of a multi-module mPKS polypeptide remains an ambitious target. The

precise mechanism of vectoral biosynthesis and the process by which polyketide intermediates are transferred between modules remain largely unexplored, with this information critical in informing future pathway engineering efforts. Recent studies redefining module boundaries appear to have positively impacted re-engineering approaches, but it too early to assess their significance with any great confidence. *In vitro* studies are also at a premium, with most investigations to date carried out *in vivo*. Expanding the scope and complexity of *in vitro* studies will allow for robust quantitative biochemical analysis, and the application of steady-state and transient kinetic methods. Finally, it is desirable to widen our knowledge of mPKS systems, with current dogma derived largely from studies of only the erythromycin and pikromycin synthases. It is unlikely that studies of these two systems alone will fully capture the peculiarities and nuances of the breadth of mPKS enzymology.

## 1.7 Biosynthetic Tailoring Enzymes

Downstream tailoring enzymes are ubiquitous in all secondary metabolite biosynthetic pathways, and are capable of performing intriguing chemical transformations. These tailoring enzymes can alter the activity and pharmacodynamic properties of the pathway product, and are often critical in ensuring the stability, solubility and bioactivity of NPs.[171] Tailoring enzymes commonly include oxygenases, reductases, halogenases, cyclases, ATs and glycosyltransferases (GTs), amongst others.

### 1.7.1 Redox Enzymes in Natural Product Biosynthesis

Redox enzymes are commonly found performing tailoring reactions in polyketide pathways. Those that catalyse the insertion of oxygen atoms into the NP scaffold are usually found as monooxygenases, which transfer one oxygen atom and reduce the second to water, although dioxygenases do exist.[172] Flavoprotein oxygenases, whose activities are reliant on FMN or FAD cofactors, have evolved to use molecular oxygen as a substrate, rather than suppressing interactions between the reduced flavin core and oxygen, as is common to most flavoproteins.[173] Monooxygenases are capable of performing a diverse set of reactions including Baeyer-Villiger oxygenation, yielding a rare carbonate moiety within the NP scaffold, for example in the case of CcsB from the fungal cytochalasin E pathway. These enzymes typically employ NADPH as a source of reducing potential, inserting an oxygen atom into a carbonyl group to form an ester. Other flavin-dependent proteins have evolved to insert halogens into NPs. This is an incredibly powerful tool in biosynthesis, frequently enhancing bioactivity, as is the case for vancomycin (Figure 1.3).[171] Of particular note is the enzyme RebH, which, along with its flavin-dependent partner RebF, catalyses the insertion of chlorine during the biosynthesis of rebeccamycin (Figure 9).[174] The RebH/RebF couple also illustrate the potential for tailoring enzymes to become industrial biocatalysts. Canonically they halogenate tryptophan, but have been successfully engineered to act upon a variety of non-natural substrates, including tryptamine[175] and arenes.[176] These enzymes can also be engineered to enhance product yield and activity. For example, the halogenase CtcP, involved in the biosynthesis of the broad spectrum antibiotic chlortetracycline, has been subjected to structure guided mutagenesis to enable elevated production of chlortetracycline by 1.7 fold *in vivo*.[177]

### 1.7.2 Radical SAM Enzymes in Natural Product Biosynthesis

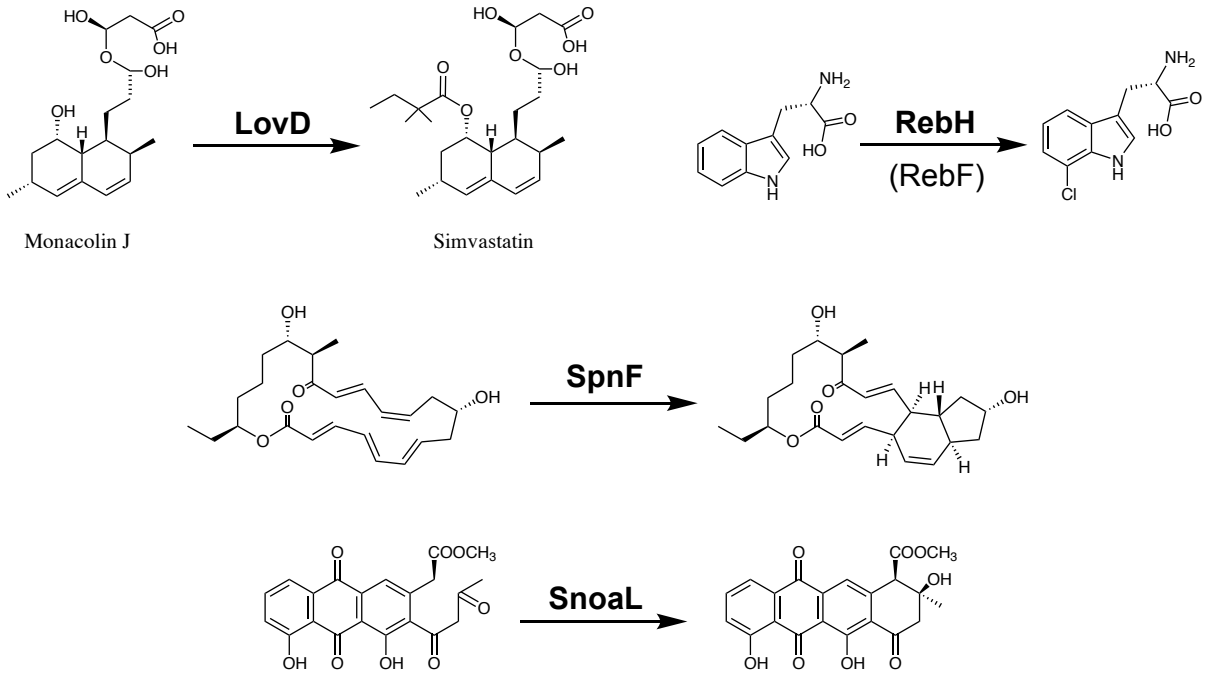
S-adenosylmethionine (SAM) dependent enzymes also perform tailoring roles as part of NP biosynthetic clusters, particularly in the synthesis of RiPPs.[178] Radical SAM enzymes can function as methyltransferases, as in the synthesis of the antibiotic anti-tumour agent thiostrepton A,[179] or as halogenases. For example, SalL catalyses chlorination of the proteasome inhibitor salinosporamide A, a marine NP isolated from *S. tropica*.[180] This enzyme has also been shown to accept bromine and iodine substrates, which can be selectively introduced to give modified salinosporamides.[181]

### 1.7.3 Engineering Transferases from Natural Product Biosynthetic Pathways

GTs and ATs from biosynthetic pathways have also been developed as industrially relevant biocatalysts.[171] Xie and Tang successfully demonstrated a single step production of simvastatin from monacolin J, a process that otherwise requires a five step synthesis, en-route to the production of this cholesterol lowering agent (Figure 1.9).[182] Researchers have also demonstrated the capacity for significant improvement in the catalytic performance of tailoring enzymes from this pathway, with engineered variants of the AT LovD exhibiting a > 11x enhancement in catalytic efficiency vs. the wild type enzyme.[183] GTs can facilitate significant modifications in NP scaffolds. These enzymes are found in a plethora of biosynthetic pathways, modifying a number of clinically relevant NPs.[184] Whilst these enzymes do not normally display substrate promiscuity, the GT involved in oleandomycin biosynthesis, OleD, which was isolated from *S. antibioticus*, has been shown capable of accepting non-natural substrates following extensive modification via directed evolution.[185] The directed evolution screen used during this study was subsequently employed to select for further engineered variants of OleD that were employed in the synthesis of multiple functionalised sugar-nucleotides. The best performing OleD variant was capable of incorporating 6 functionalised sugars and 5 nucleoside diphosphate acceptors, yielding a library of 30 distinct sugar-nucleotides.[186] Similarly, Parajuli *et al*. have used promiscuous GTs to diversify existing NPs, aiming to increase solubility and bioactivity. In this study the GT YijiC was shown to be capable of producing 6 non-natural glycosylated analogues of epothilone A.[187]

### 1.7.4 Standalone Cyclases in Natural Product Biosynthesis

Although TE domains within mPKS generally considered to be the primary driver of product cyclisation during the biosynthesis of polyketide NPs, many biosynthetic pathways also incorporate stand-alone cyclases in their compliment of tailoring enzymes. Terpene biosynthetic clusters contain a multitude of cyclases that are amongst the most well understood.[188] Cyclases in PKS and NRPS systems are less well characterised, though there have been a number of interesting candidates identified. SpnF, a cyclase involved the biosynthesis of the insecticide spinosyn A, has been shown to catalyse a [4 + 2] cycloaddition reaction, producing a cyclohexene product, in a process analogous to a formal Diels-Alder (DA) reaction (Figure 1.9).[189,190] Other cyclases, such as the anthracycline cyclase SnoaL, use an aldol condensation reaction to facilitate the closure of ring systems.[191]




Monacolin J Simvastatin

**Figure 1.9 Selected Tailoring Reactions Performed During the Biosynthesis of NPs.**

Undoubtedly, there is a significantly greater diversity in the structures and functions of NP pathway tailoring enzymes than is observed in core NP pathway synthases. They represent viable targets for rational engineering once fully characterised and are often more tractable to work with in heterologous hosts due to their modest size. Substrate specificity can be altered,

with particularly impressive results achieved using high-throughput directed evolution techniques. They do, however, lack the inherent modularity and potential predictability of mPKSs. The study and engineering of these biocatalysts is of high priority and has the potential to unlock new biosynthetic routes enabling access to a diversity of novel natural products.

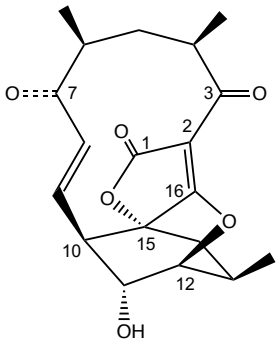
## 1.8 Abyssomicin C

### 1.8.1 The Isolation of Abyssomicin C

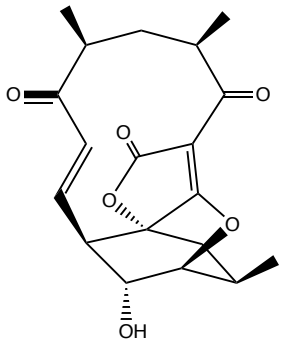
As part of an ongoing effort to address the antibiotic crisis through the discovery of new NPs from unusual environments, in 2004 Bister *et al*. reported the discovery and isolation of abyssomicins B, C, and D from *Verrucosispora* sp. AB 18-032 (Figure 1.10).[192] Recently, however, the formal classification of this actinomycete strain as *Verrucosispora maris* was revised to *Micromonospora maris,* as part of a major reclassification of the actinomycetes in their entirety by Göker, Goodfellow and co-workers.[193] *M. maris* was first isolated from a sediment sample obtained at a depth of almost 300 m off the coast of Japan. During the isolation and testing of these three compounds, abyssomicin C was found to represent the major product, as well as being the most bioactive of the three, exhibiting potent bioactivity against a variety of Gram-positive bacteria, including methicillin-resistant *Staphylococcus aureus* (MRSA) and vancomycin resistant *S. aureus* (VRSA) isolates.[194] The potential value of abyssomicin C as an antibiotic lead was further enhanced following the establishment of its mode of action, which was found to be *via* inhibition of the key *para*-aminobenzoic acid (*p*ABA) biosynthesis enzyme PabB. *p*ABA is a precursor required for the biosynthesis of tetrahydrofolate, itself essential for the production of aromatic amino acids *via* the folate biosynthetic pathway.

### 1.8.2 Atrop-Abyssomicin C

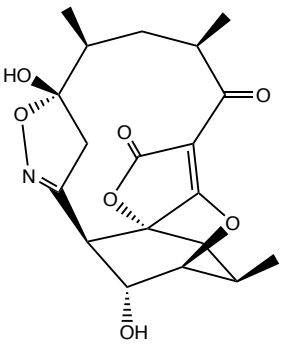
Following the isolation of abyssomicin C, a stereoisomer, atrop-abyssomicin C, was synthesised by Nicolaou and Harrison, and shown to have an increased potency against MRSA (Figure 1.10).[195] It was also demonstrated that incubation of abyssomicin C with a strong acid, including those frequently used in high performance liquid chromatography (HPLC) separation, would catalyse the conversion of atrop-abyssomicin C to abyssomicin C. Careful re-analysis of *M. maris* extracts confirmed atrop-abyssomicin C as the principle compound produced, and with the most potent inhibitory effect.[196,197] There have been difficulties, however, in precisely assigning this atrop-isomerism. Various sources represent the isomerism between the ketone at position 7 and the end of the alkene system at position 10 differently.[197–201] The depiction of **1** in Figure 1.10 is the most commonly used representation.



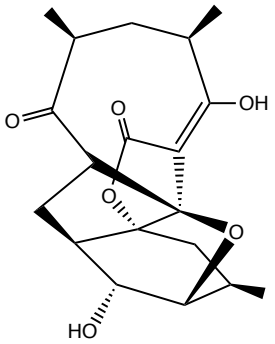
Atrop-abyssomicin C (**1**)



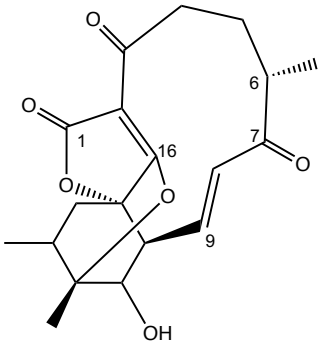
Abyssomicin C (**2**)



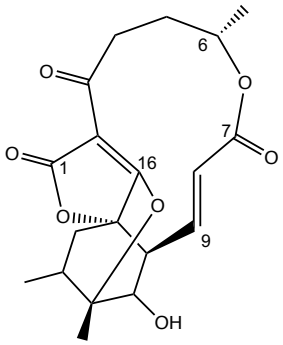
Abyssomicin B



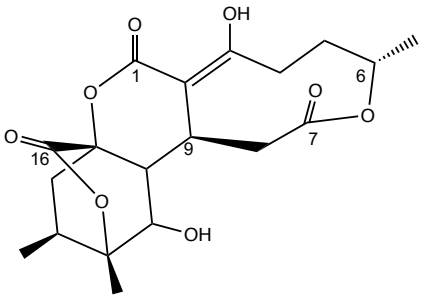
Abyssomicin D



Abyssomicin 2 (**3**)



Neoabyssomicin B **(4)**



Neoabyssomicin A

**Figure 1.10 Chemical Structures of Key Abyssomicin and Neoabyssomicin NPs.** Numbering is consistent for abyssomicins B – D and consistent between abyssomicin 2, neoabyssomicin A and neoabyssomicin B.

### 1.8.3 The Folate Biosynthetic Pathway

The folate biosynthetic pathway is comprised of the shikimate, chorismate and folic acid (SCF) pathways (Figure 1.11). *p*ABA is biosynthesised from chorismate, which is in turn generated through the shikimate pathway. *p*ABA biosynthesis requires two enzymes, 4-amino-4-deoxychorismate synthase (ADCS), which is comprised of the PabA and PabB monomers, and 4-amino-4-deoxychorismate lyase (ADCL), also known as PabC. *p*ABA is, in turn, required for folic acid biosynthesis, where it is condensed to pteridine by the enzyme dihydropteroate synthase (DHPS). Synthesis of folate *via* this well-studied pathway is ubiquitous in prokaryotes and some lesser eukaryotes, but is not present in higher eukaryotes, including humans, who obtain folate through consumed nutrients.[202] The SCF pathway has therefore long been considered an attractive target for inhibition by antimicrobials, due to its essential nature in microbes and anticipated reduced off-target effects in humans. The enormously successful chemical pesticide glyphosate is itself an inhibitor of the shikimate pathway and the sulphonamides represent clinically approved inhibitors of the folic acid biosynthetic pathway. The antibiotic trimethoprim, another example of a successful fully synthetic antibiotic,[203]binds to dihydrofolate reductase, preventing the production of tetrahydrofolate (THF). There was, however, until the identification of the abyssomicins, no known NP based inhibitor of folate biosynthesis.[201] A combination of pathway validation by the two synthetically derived drugs, along with the emergence of widespread resistance to sulphonamides[204,205] and trimethoprim,[206] further enhanced the attractiveness of this pathway as an antimicrobial target. The sulphonamides act as a competitive inhibitor of the enzyme DHPS, mimicking the structure of *p*ABA. A common mode of sulphonamide resistance is to increase the synthesis of *p*ABA, outcompeting the sulphonamide inhibitor. This meant that the identification of atrop-abyssomicin C as an inhibitor of ADCS, specifically PabB, *via* a Michael addition reaction to an active site cysteine (Figure 1.12), represented the identification of a novel mode of action with the potential to complement sulphonamide based therapies, whilst circumventing established resistance mechanisms.[196,201] The Michael addition reaction of abyssomicin C with PabB had been predicted based on the chemical structure of this compound, as the inactive abyssomicins B and D lack the Michael system adjacent to the ketone group. This means that these compounds, which are worse Michael acceptors, are unable to bind irreversibly to PabB.

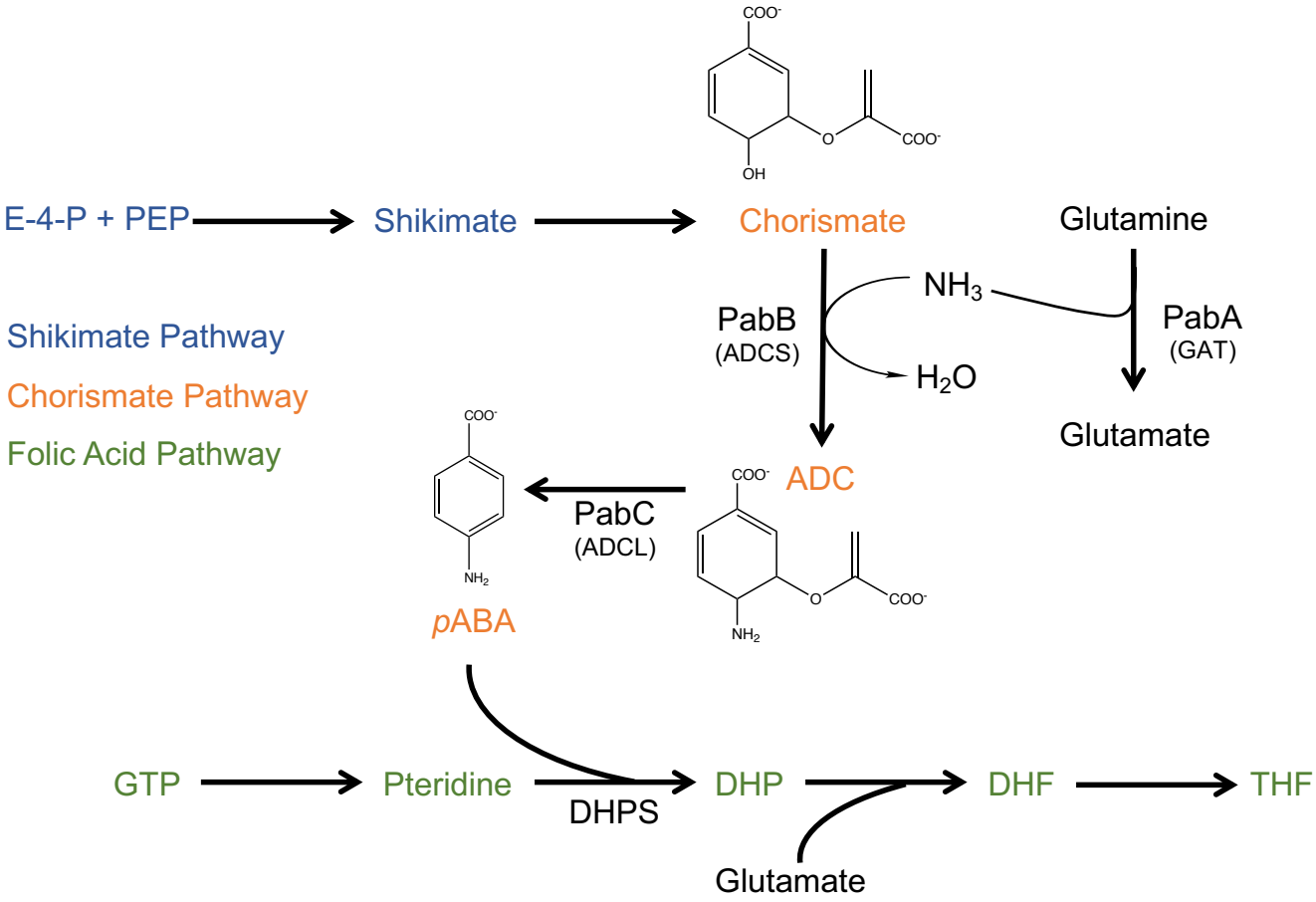



**Figure 1.11 Schematic representation of folate biosynthesis through the SCF pathways.** ADC, amino-dechorismate; ADCL, amino-dechorismate lyase; ADCS, amino-dechorismate synthase; DHF, dihydrofolate; DHP, dihydropteroate; E-4-P, erythrose-4-phosphate; GAT, glutamine amidotransferase; GTP, guanosine-5'-triphosphate; PEP, phosphoenolpyruvate; THF, tetrahydrofolate.

### 1.8.4 Abyssomicin Bioactivities

Abyssomicin C and its atrop-isomer are both defined as small spirotetronates. The small spirotetronates consist of a core, comprising a tetronic acid ring spiro-linked to cyclohexene, embedded in an 11 carbon macrocycle. There may also be additional ring structures present.[198] Within the small spirotetronates there are a number of close analogues of **2** with reported bioactivities.[201] Atrop-abyssomicin C has also been shown to display potent bioactivity against tuberculosis (TB) strains, as has the analogue abyssomicin J.[207,208] Abyssomicin J retains activity against TB isolates despite having lost the crucial Michael system. Instead this dimer of abyssomicins acts as a prodrug, undergoing oxidative activation to yield atrop-abyssomicin C.

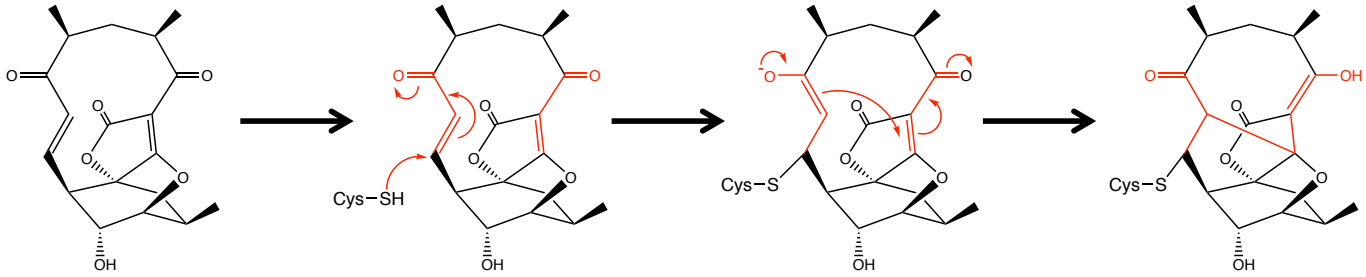



**Figure 1.12 Mechanism of Action by Abyssomicin C and its Atrop-Isomer.** Abyssomicin C and atrop-abyssomicin C inhibit folate biosynthesis through irreversible binding to the active site cysteine of PabB. This occurs due to the two Michael acceptor systems present in 1 and 2 (coloured red). Figure adapted from Savic.[200]

The Neoabyssomicins A & B were recently isolated from *Streptomyces koyangenesis*, another marine bacterial strain, and were shown to possess a dilactone bridge, differentiating them from previously reported abyssomicins (Figure 1.10).[209,210] Although these intriguing scaffolds do not yet have any assigned therapeutic uses, the precursor to neoabyssomicin B, abyssomicin 2, has been shown to re-activate latent human immunodeficiency virus (HIV).[211] This unusual activity is of interest to those seeking to cure HIV-infected patients, enabling the virus to be identified both by the immune system and by other retroviral drugs.[212] In January 2020, Zhang *et al*. reported the isolation of abyssomicin Y and its analogues. This compound was shown to have potent antiviral activity, specifically against influenza type A.[213]

These recent insights into abyssomicins lacking a Michael system indicate the potential of this core NP scaffold as a therapeutic lead. Unfortunately, abyssomicin C has been shown to exhibit cytotoxic effects on human cell lines, precluding clinical use. In an effort to address this issue, atrop-O-benzyl-desmethylabyssomicin C, featuring a protective benzene group and lacking the two methyl groups protruding from the macrocyclic ring, was synthesised by the Saicic group.[214] This molecule exhibits a much lower cytotoxic effect on human cells, but retains activity against clinical isolates of MRSA, giving hope that the cytotoxic mechanism of action differs from the Michael addition reaction necessary to inhibit PabB. Further reduction of atrop-abyssomicin C, however, has not been successful, as any molecules lacking the

bicyclic ring system produced using a synthetic Diels-Alder reaction have been shown to be inactive.[215]

### 1.8.5 Synthesis of Atrop-Abyssomicin C

The therapeutic potential and novel mode of action of atrop-abyssomicin has resulted in this compound being considered an attractive target for total synthesis, with a number of different groups having now developed syntheses of this compound and its analogues. The first synthesis completed was that of Sorensen and co-workers, and was quickly followed by an alternative synthesis completed by Nicolaou and Harrison.[195,216] Both of these reported routes made use of Diels-Alder (DA) reactions to form the central chiral atom. Further synthetic attempts included the synthesis of (−)-atrop–abyssomicin C by Bihelovic *et al*., resulting in an enantioselective product, albeit at only 1.8% yield.[199] The earlier, non-enantioselective syntheses occurred at similar yields, although with fewer than the 21 steps used by Bihelovic *et al*. Motivated by these studies  attention began to turn to establishing the biosynthetic origins of the abyssomicins and exploring possibilities for the development of semi-(bio)synthetic, analogues.

## 1.9 Abyssomicin Biosynthesis

### 1.9.1 Characterising the Abyssomicin C Biosynthetic Gene Cluster

The abyssomicins are not only of interest due to their potential clinical value, but also for their unique and varied chemical structures that appear to incorporate many unusual biosynthetic steps.[198] Upon initial isolation, abyssomicin C was formally classified as a polyketide antibiotic,[192] however, it took until 2011 for a formal investigation into the biosynthesis of abyssomicin C to take place. By this time the mechanism of action for abyssomicin C had been determined,[196] and three total syntheses had been reported, though none with a final yield above 4%.[216–219] A rigorous growth condition assessment of *M. maris* had not been undertaken, and bacteria from the now redundant genus *Verrucosispora* had never before been successfully genetically manipulated in the laboratory. Süssmuth and co-workers successfully combined interrogation of the BGC encoding abyssomicin C with $C^{13}$ labelled substrate feeding studies.[220] These revealed that abyssomicin C was constructed from 5 acetate units, two propionate units and a glucose based 2-carbon molecule. This fits well with the genetic profiling, which identified the Type I mPKS genes *aby*B1, *aby*B2 and *aby*B3 as containing 7 modules, predicted to insert the 5 malonyl-CoA and 2 methylmalonyl-CoA starter and extender units required to form the linear carbon chain. Module 5 was predicted to contain inactive KR and DH domains in order to maintain collinearity, and genetic analysis determined that inactivating mutations were present in their active sites.

Downstream of the mPKS modules, a variety of tailoring enzymes were identified. Inactivation of any of the enzymes AbyA1 – AbyA4 proved to be catastrophic for the production of abyssomicin C, and, for these reasons, as well as genetic comparisons to the Chlorothricin and Kijanimicin BGCs,[221,222] they were predicted to be involved in the formation of the tetronate moiety, inserting a glycolytic pathway-generated 2-C molecule. In addition to this, a number of genes hypothesised to be involved in the export of the antibiotic molecule (*aby*F1 – *aby*F4) were identified, along with a cytochrome P450 system (*aby*V, *aby*W and *aby*Z), regulators and monooxygenases. Notably, however, Gottardi *et al*. were unable to identify any enzyme thought capable of catalysing the intramolecular DA reaction necessary to cyclise the linear polyketide chain **5**.

### 1.9.2 Subsequent Characterisation of Individual Enzymes in the Abyssomicin C Biosynthetic Gene Cluster

Following the first analysis of the abyssomicin C BGC, further studies have elucidated the function, structure and mechanisms by which the mPKS gene products and associated tailoring enzymes assemble the active compound, atrop-abyssomicin C. An up-to-date overview of this pathway is shown in Figure 1.13. The entire BGC is ~57 kb long, of which *aby*B1 accounts for > 17 kb (around 30 %), *aby*B2 > 11 kb and *aby*B3 ~3 kb. Whilst the three mPKS genes comprise > 55 % of the BGC, there are a minimum of 24 further putative enzymes involved in the biosynthesis and export of **2**, not all of which have been fully characterised to date. Due to many of the issues already discussed the mPKS proteins encoded, AbyB1-3, have only recently been subjected to any level of *in vitro* characterisation.[223] Heterologous expression of AbyB3 has been achieved, the details of which are outlined in the results section of this thesis. While not comprehensively studied, the route to the tetronate moiety has been elucidated through comparative genomic studies. AbyA3 has been identified as a standalone ACP, responsible for reducing glycerate in conjunction with AbyA2, and transferring the product to the polyketide backbone. AbyA1 catalyses C-C bond formation, simultaneously forming the tetronate ring and liberating the AbyB3 tethered polyketide chain.[198,224] Once the tetronate ring has installed, AbyA4 is hypothesised to acetylate the linear chain *via* attack at the hydroxyl moiety. This prepares the chain for modification by an acetate lyase, AbyA5. This class of enzyme is unique to the formation of spirotetronate/spirotetramate NPs, creating the dienophile necessary for DA cyclisation. AbyA5 has recently been shown to be a stereo-specific catalyst, efficiently eliminating acetate without the production of hydrolysis products in the natural system.[225] This linear chain, **5**, now contains the diene system in the hydrocarbon tail and the dienophile, created by AbyA5, both of which are necessary for an enzyme-catalysed [4 + 2] cyclisation by a Diels-Alderase (DAase). This enzyme, AbyU, was reported in 2015 by Byrne *et al*., and was the first reported natural enzyme specifically evolved to catalyse a DA reaction.[224] AbyU, like all spirotetronate cyclases, does not share any sequence similarity with putative DAases from other systems and hence was not included in the original BGC annotation by Gottardi *et al*. It was shown to catalyse the [4 + 2] cycloaddition reaction *via* a concerted asynchronous DA reaction. This will be further elaborated in Chapter 5. The final steps in the conversion of the cyclised *pre*-abyssomicin **6** are not yet fully understood. It is anticipated that a combination of AbyV and E create an epoxide group at the expense of the hexene double bond. This highly reactive intermediate is then quickly processed to become **2**. Subsequent ring opening and formation of

a heteromeric, bicyclic, 6-membered ring system is probably catalysed and controlled by further enzymes, such as AbyV E and Z.[198]

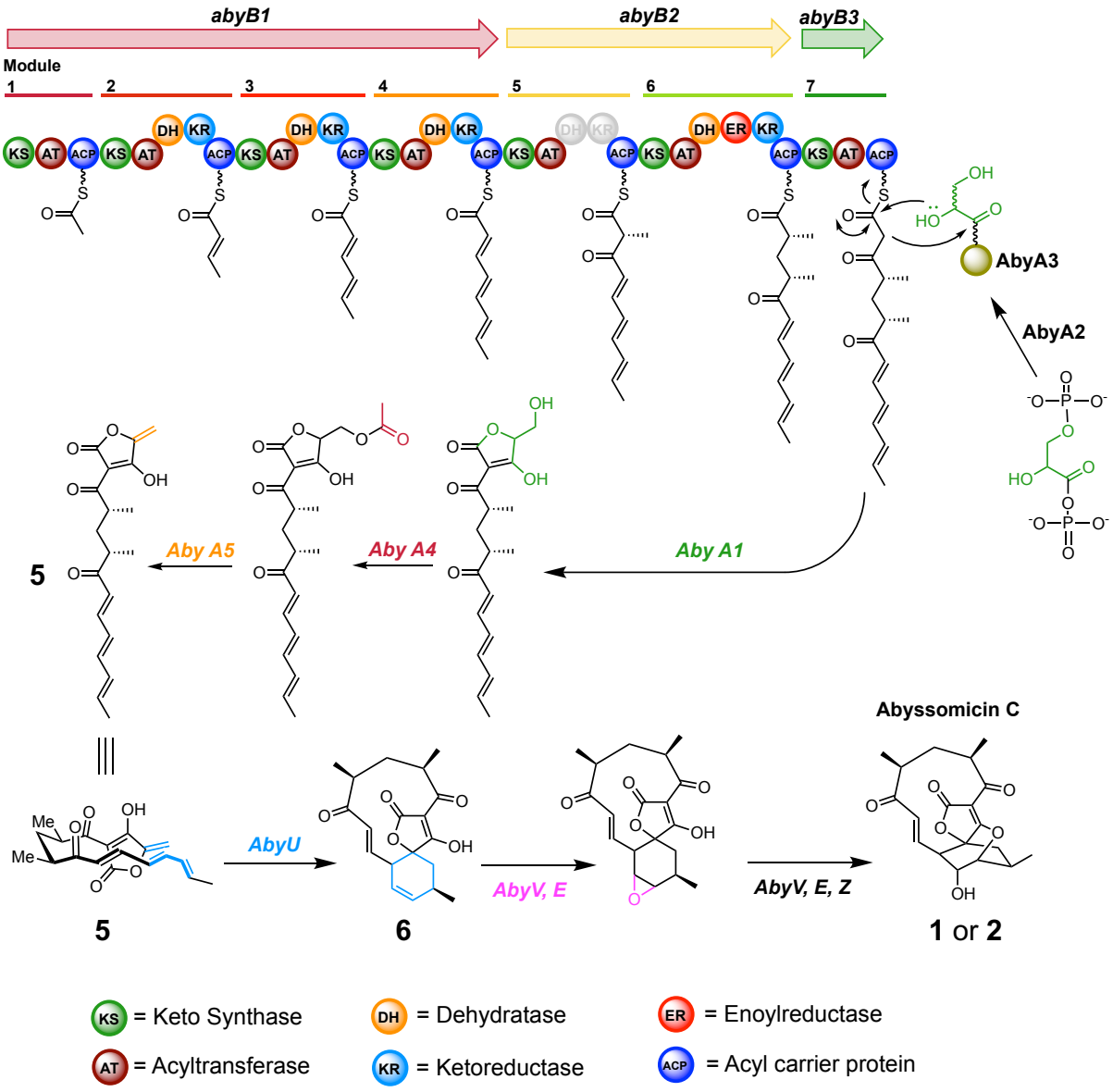



**Figure 1.13 A Schematic of the Abyssomicin C Biosynthetic Pathway in *M. maris*.**

### 1.9.3 Neoabyssomicin Biosynthesis

Abyssomicin 2 (**4**), a close analogue of abyssomicin C, is biosynthesised by a highly analogous enzymatic pathway (Figure 1.14). Further tailoring steps are then possible to produce abyssomicin 4, neoabyssomicin A and neoabyssomicin B.

Similarly to the abyssomicin C pathway, the neoabyssomicin pathway comprises 3 multi-modular mPKS polypeptides, housing 7 modules. 4 of these are located on AbmB1, 2 on AbmB2 and a single minimal module resides on AbmB3. Each module performs the same level of reduction on the growing polyketide chain as in the biosynthesis of abyssomicin C. There are, however, distinct differences in the substrate specificity of modules 2 and 6. Module 2 loads methylmalonyl-CoA instead of malonyl-CoA, and module 6 loads malonyl-CoA as opposed to methylmalonyl-CoA. These two small changes in chain decoration, in combination with the alternative stereoisomer generated during DA catalysed cyclisation, translate to a modified bioactivity, with abyssomicin 2 exhibiting potent antiviral rather than antibiotic activity. The product of the PKS assembly line is similarly modified by the addition of a tetronate head group, installed by the enzymes AbmA1 – AbmA5 (analogous to AbyA1-A5). AbmU is then hypothesised to catalyse the intramolecular DA reaction, yielding the cyclised isomer of **7**. The epoxide necessary to create the final ring system is formed through action by AbmV and E, before formation of the tetracyclic **3** by AbmV and AbmZ.[201,226] In order to convert **3** into **4**, a Baeyer-Villiger reaction takes place, inserting an oxygen atom into the macrocycle. AbmE2 and AbmZ have recently been shown to catalyse this interesting reaction, acting in concert, with AbmZ functioning as a reductase, allowing AbmE2, a luciferase-like monooxygenase, to perform the oxygenation step.[227] Together these constitute a Type II Bayer-Villiger monooxygenase system, another example of a rare and poorly understudied biocatalytic process.

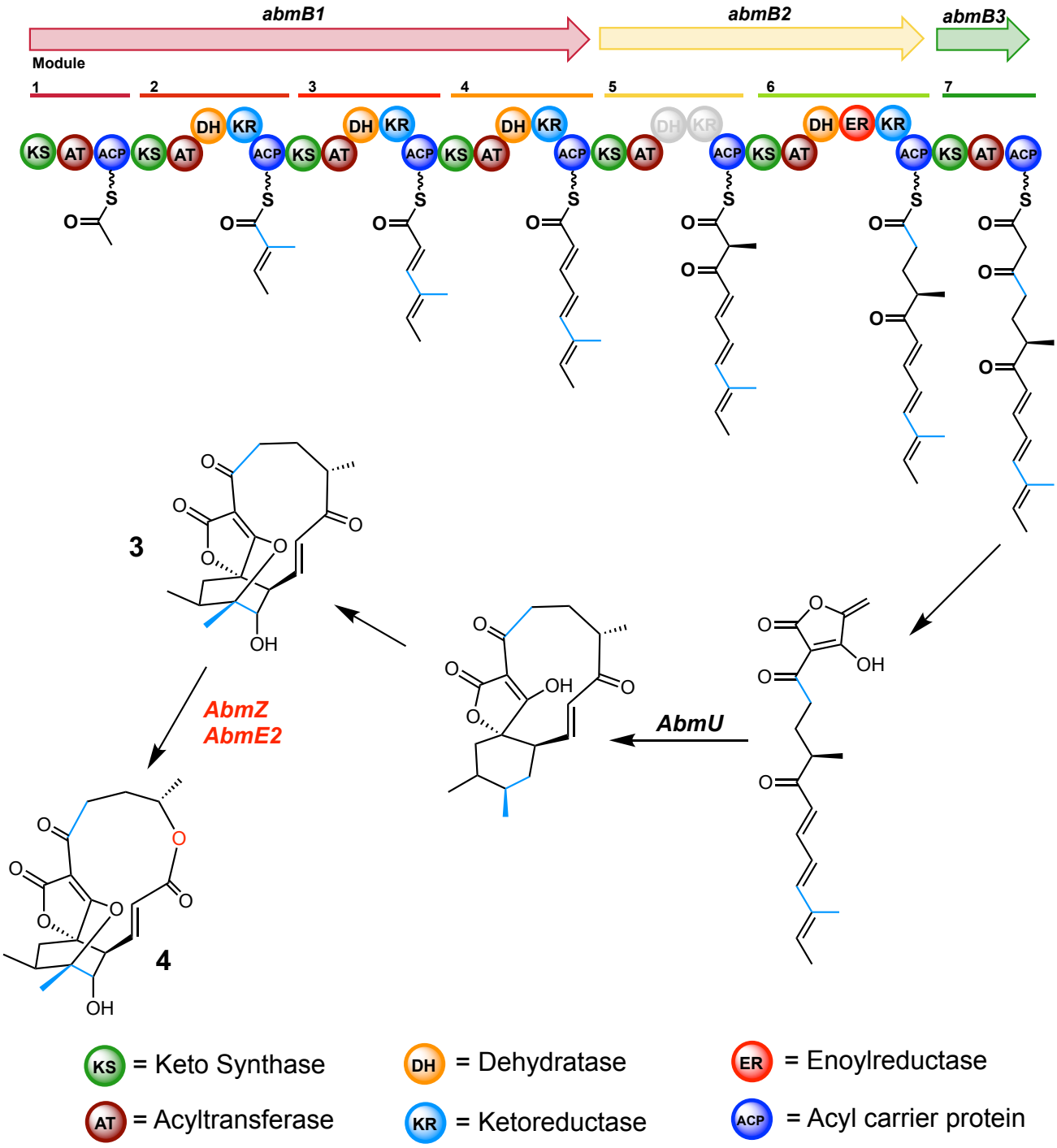



**Figure 1.14 A Schematic of the Neoabyssomicin Biosynthetic Pathway from *S. ko-yangenesis*.** The only mPKS difference in comparison to the abyssomicin C biosynthetic pathway is in the loading specificity for methylmalonyl-CoA at module 2 and malonyl-CoA at module 6. The resulting change in chain decoration is coloured blue throughout.

## 1.10 Research Aims

NP biosynthetic pathways are rich sources of novel chemistries and exploitable enzyme biocatalysts. The abyssomicin C biosynthetic pathway represents an opportunity to further our understanding of Type I PKSs, their products, and the suite of tailoring enzymes that are employed to create the functional biomolecule. Progress made in addressing the significant gaps in our understanding of these pathways will facilitate future engineering attempts to exploit and improve their products, potentially leading to novel, clinically relevant, compounds. This thesis aims to address these gaps in our understanding through the interrogation of two critical enzyme classes represented within the abyssomicin C biosynthetic pathway: The Type I PKS assembly lines and the DAase AbyU. Specifically, studies have been performed to develop cloning and expression protocols for the mPKS polypeptides AbyB2 and AbyB3. Subsequently, AbyB3 was subjected to a structural characterisation using complementary X-ray crystallography and cryo-EM methods. AbyU has been mechanistically interrogated using transient techniques to determine a deceptively complex reaction mechanism, with further understanding developed through the application of high-pressure structural and functional analyses. The studies performed herein on AbyU represent the first mechanistic characterisation of its kind for a [4 + 2] cyclase or enzyme displaying a β-barrel architecture, whilst the structural characterisation of AbyB3 presents a significant step towards reconciling the conflicting theories of mPKS modular architectures. As such the work has potential to significantly improve our understanding of these systems and impact on future engineering attempts.

# Chapter 2

# General Methods

## 2.1 Bioinformatic Analyses

The following software programmes and online tools were used to conduct bioinformatic analyses of target nucleic acid and protein sequences, and for primer design and gene codon optimisation.

- Benchling: https://benchling.com/
- Clustal Omega Multi-Sequence Alignment: https://www.ebi.ac.uk/Tools/msa/clustalo/
- DNA-DNA Hybridization: http://ggdc.dsmz.de/home.php
- Emboss Needle Pairwise Sequence Alignment: https://www.ebi.ac.uk/Tools/psa/emboss_needle/
- Expasy Translate: https://web.expasy.org/translate/
- *M. maris* Genome: https://www.ncbi.nlm.nih.gov/nuccore/NC_015434.1
- Prot pi Bioinformatics Calculator: https://www.protpi.ch/
- ThermoFisher GeneArt Primer and Construct Design Tool: https://www.thermofisher.com/order/oligoDesigner/
- ThermoFisher GeneArt Project Manager: https://www.thermofisher.com/order/geneart-genes/projectmgmt

## 2.2 Microbiological Techniques

### 2.2.1 Polymerase Chain Reaction

DNA sequences encoding target enzymes were either ordered from ThermoFisher GeneArt as synthetic genes, or directly amplified from *M. maris* genomic DNA. Gene amplification was performed using polymerase chain reaction (PCR). PCR reaction mixes were established as outlined in Table 2.1. In all cases master mix solutions were employed, comprising pre-mixed dNTPs, buffer, thermostable DNA polymerase enzyme and hot start antibody, included to minimise non-specific amplification. Unless otherwise specified, CloneAmp HiFi Master Mix (Takara Bio) was used for PCR reactions, employing the amplification conditions listed in Table 2.2.

| Component | Volume (µL) | Concentration |
|---|---|---|
| Total Volume | 25 / 50 | - |
| 2x Master Mix | 12.5 / 25 | - |
| Template DNA | 1 / 2 | 1 ng/µL |
| Forward Primer | 0.75 / 1.5 | 10 pMol |
| Reverse Primer | 0.75 / 1.5 | 10 pMol |
| $ddH_2O$ | 10 / 20 | - |

**Table 2.1 Typical PCR Reaction Mix Composition.**

| Temperature (°C) | Time (s) | Cycles |
|---|---|---|
| 98 | 10 | 1 |
| 98 | 5 | 30 |
| 62 | 30 | |
| 72 | 30 | |
| 72 | 300 | 1 |

**Table 2.2 Typical PCR Amplification Conditions.**

### 2.2.2 Agarose Gel Electrophoresis

PCR products were visualised using agarose gel electrophoresis. 1 % agarose gels were made by dissolving 0.5 g of agarose powder in 50 mL of Tris/acetate/EDTA (TAE) buffer (40 mM Tris base, 20 mM acetic acid and 1 mM EDTA) supplemented with 5 μL of SYBR Safe™ dye (ThermoFisher, 10,000 x concentration). PCR reactions were mixed with 6x concentrate DNA gel loading dye (ThermoFisher) prior to gel loading and electrophoresis. 2 μL of each PCR reaction was used for sample analysis, with entire reaction mixes (50 μL) used for gel extraction. Gels were electrophorised at 120 V for 25 minutes in a TAE buffer. Gels were visualised under UV light to assess sample quality and quantity. DNA extractions from agarose gels were performed using the QIAquick PCR purification kit (Qiagen), following the manufacturer's instructions.

### 2.2.3 Plasmid Preparation

Purified PCR products were inserted into destination plasmids (Appendix Table A2) *via* homologous recombination, employing the In-Fusion cloning kit (Takara Bio), as per the manufacturer's instructions. 5 μL samples of each recombination mix were transformed into chemically competent *E. coli* DH5α Stellar cells (Clontech) by 45 s heat shock at 42 °C for 50 seconds. Following heat shock, cells were subjected to a 45 min outgrowth in LB broth before transfer to agar plates. Agar plates were prepared by dissolving 12 g of granular agar in 1 L LB broth, followed by supplementation with either 100 μg/mL carbenicillin (Carb) or 50 μg/mL kanamycin (Kan), dependent on the antibiotic selection marker present in the host plasmid. Plates were incubated at 37 °C for 18h before a single colony was picked and used to inoculate 10 mL LB broth supplemented with appropriate antibiotic as above. Cultures were grown for 18h at 37 °C with shaking, following which cells were harvested by centrifugation and plasmids were isolated using the QIAprep plasmid miniprep kit (Qiagen), following the manufacturer's instructions. 1 μL of each purified plasmid was used to transform 20 μL of a chemically competent *E. coli* expression strain, typically *E. coli* BL21 (DE3) (NEB) or *E. coli* BL21 (DE3) BAP1 cells (prepared in-house, see 2.2.4) via the heat shock method. Transformed cells were plated onto LB-agar plates as outlined above in preparation for protein expression studies.

### 2.2.4 Preparation of Chemically Competent *E. coli* Cells

*E. coli* cells of the desired strain were plated onto LB-agar plates and grown overnight at 37 °C. Single colonies were picked and grown in 10 mL LB broth without antibiotic. 0.5 mL was used to inoculate 50 mL of fresh LB broth, which was subsequently grown with vigorous shaking to an $OD_{600\ nm}$ of 0.4 – 0.6. Cells were subsequently harvested by centrifugation at 4 °C. All subsequent preparation steps were conducted in a constant temperature room at 4 °C. Cell pellets were gently resuspended in 5 mL TBF I buffer (30 mM potassium acetate, 100 mM rubidium chloride, 50 mM manganese chloride, 10 mM calcium chloride, 15% glycerol, pH 5.8) and left on ice for 10 minutes. Cell suspensions were then centrifuged at 1500 x g and supernatants removed. Resulting cell pellets were gently resuspended in TBF II buffer (10 mM MOPS, 10 mM rubidium chloride, 75 mM calcium chloride, 30 % glycerol, pH 6.5), before 50 µL aliquots were removed and flash frozen in liquid nitrogen.

## 2.3 Protein Expression, Purification and Analysis

### 2.3.1 Small Scale Expression Testing

0.5 mL of *pre*-transformed *E. coli* cell culture was used to inoculate 50 mL of growth media, either LB, Terrific Broth (TB) or autoinduction TB (TBAI) broth, supplemented with appropriate antibiotic. Cultures were grown to at 37 °C with shaking at 180 rpm to an $OD_{600\ nm}$ of either 0.4-0.6 for cultures grown in LB, or ~1 for those grown in TB. Once the required optical density had been reached, cultures were transferred to a shaking incubator at 20 °C, followed by induction with up to 1 mM IPTG. The addition of IPTG was not required for cultures grown in autoinduction media. Following induced growth for 18h, aliquots of cell cultures were removed, cells harvested by centrifugation, and samples visualised by SDS-Polyacrylamide gel electrophoresis (SDS-PAGE) as outlined below.

### 2.3.1 Large Scale Protein Expression

*E. coli* transformants were grown on agar plates supplemented with appropriate antibiotic. Single colonies from these plates were used to inoculate 10 mL of LB broth supplemented with appropriate antibiotic. Cultures were grown overnight to turbidity at 37 °C, with shaking at 180 rpm. These starter cultures were subsequently used as a 1 % inoculant for 1 L of the required growth media, LB broth for AbyU or TB for AbyB3 and AbyB3ΔACP. All growth media were supplemented with 100 μg/mL Carb. Cultures were grown at 37 °C with shaking at 180 rpm to either an $OD_{600\ nm}$ of 0.4-0.6 for LB based cultures or ~1 for TB based cultures. Once the required OD had been reached, cultures were induced (1 mM IPTG for AbyU, 0.2 mM IPTG for AbyB3 and AbyB3ΔACP) and grown at 20 °C with shaking at 200 rpm for 18 h prior to being harvested by centrifugation. Resulting cell pellets were stored at -70 °C prior to lysis.

### 2.3.2 Protein Purification

Cell pellets were thawed and resuspended in His Load buffer (20 mM Tris-HCl, 150 mM Sodium Chloride, 20 mM Imidazole, pH 8). Cell suspensions were homogenised and passed through a cell disruptor (Constant Systems Ltd), with resulting cell lysates clarified by centrifugation at 39,000 x g for either 30 mins (LB based cultures) or 45 mins (TB based cultures).

Supernatants were removed and passed through a 5 mL HiTrap HP column (GE life sciences) pre-charged with nickel. The column was washed with Load buffer before bound protein was eluted over 40 mL using a linear gradient of 0-50% Load to Elute buffer (150 mM NaCl, 20 mM Tris-HCl, 1 M imidazole, pH 8), with 1.5 mL fractions collected throughout. The absorbance of the column eluent was monitored throughout at 280 nm, with all peak fractions subjected to SDS-PAGE analysis. Eluted fractions containing the protein of interest at > 75 % purity were pooled and concentrated to 2 mL. Concentrated proteins were subsequently loaded onto either a GE Healthcare 16/60 S75 column for AbyU, or a GE Healthcare 16/60 S200 column for both AbyB3 and AbyB3ΔACP. Both columns were *pre*-equilibrated in size exclusion chromatography (SEC) buffer (150 mM NaCl, 20 mM Tris-HCl, pH 7.5) prior to sample loading. This buffer system was also used for protein elution, which was performed at 1 mL/min, with the absorbance of the column eluent monitored throughout at 280 nm. All peak fractions were subjected to SDS-PAGE analysis to assess protein purity and quantity. Protein fractions determined to be of > 95 % purity were pooled and concentrated using a Vivaspin 20 mL centrifugal concentrator (Sartorius). A 10 kDa molecular weight (MW) cut-off was used for AbyU and a 50 kDa MW cut-off used for AbyB3 and AbyB3ΔACP.

### 2.3.3 SDS-PAGE

Proteins were resolved using either precast 4-20 % acrylamide gels (BioRad or NuSep), or by employing in-house gels comprising a 7, 12 or 15% resolving gel as detailed in Table 2.3.

| Solution | 7 % Resolving gel (µL) | 12 % Resolving gel (µL) | 15 % Resolving gel (µL) | 5 % Stacking gel (µL) |
|---|---|---|---|---|
| Acrylamide 40 % | 175 | 300 | 375 | 125 |
| Tris-HCl 1.5 M pH 8.8 | 250 | 250 | 250 | ----- |
| Tris-HCl 0.5 M pH 6.8 | ----- | ----- | ------ | 250 |
| Water | 555 | 430 | 355 | 605 |
| AMPS 10 % | 8 | 8 | 8 | 12 |
| TEMED | 0.7 | 0.7 | 0.7 | 1.4 |

**Table 2.3 Polyacrylamide Gel Preparation.** Each column describes the quantities necessary to make 1 mL of gel mix. Typically, 2 gels were cast simultaneously using a combined 8 mL resolving solution and 2 mL stacking solution. AMPS, ammonium persulphate; TEMED, Tetramethylethylenediamine.

For sample analysis 2 µL from each fraction to be analysed was added to 5 µL SDS-PAGE loading dye (4% SDS, 125 mM Tris-HCl pH 6.8, 20 % Glycerol, 0.004 % Bromophenol Blue, 0.02% 2-mercaptoethanol) and denatured by heating at 95 °C for 5 minutes. Samples were subsequently loaded into the wells of the gel and subject to electrophoresis at 220 V for 35 minutes. Gels were then stained using Coomassie blue solution (40% MeOH, 7% Acetic acid, 0.25% Coomassie Brilliant Blue R) and de-stained using a 40% MeOH, 7% acetic acid solution.

# Chapter 3

# Heterologous Production of Recombinant Polyketide Synthase Modules and Multi-Modular Polypeptides

## 3.1 Introduction

### 3.1.1 Heterologous Expression and Model Organisms

The ability to express individual genes and entire gene clusters within hosts aside from their native producers has been a cornerstone of biotechnology and medicine for decades. This technology allows us to study proteins, enzymes and pathways of interest within tractable and well-characterised host organisms.[228] Within the plant world, *Arabidopsis thaliana* has become the model organism of choice,[229,230] affording a proxy for basic research into plant biology, structure and function. It has also enabled extensive research into the natural biology of crops, as well as their genetic modification to increase nutritious value or resilience.[231–233] In the animal and particularly mammalian world, the development of well characterised human cell lines and animal models, frequently mice,[234,235] have proven similarly transformative in establishing the fundamentals of animal biology[236] and in the identification of therapeutic candidates prior to the commencement of clinical trials.[237,238] *Saccharomyces cerevisiae* have been developed as the simplest of eukaryotic model organisms,[239] whilst *E. coli* is the undisputed model organism of the prokaryotic world.[240]

There are several prerequisites for the establishment of a viable and universal model organism. They must be cheap to maintain (relatively speaking), quick to grow and genetically tractable. Crucially, they must also provide a viable imitation of the natural producer or molecular target for drug efficacy trials.[241] This means that it is preferential to use as simple a model as possible for the work to be undertaken, whilst preserving key characteristics necessary for comprehensive analysis. For investigating complex therapeutic responses or disease

characteristics in humans, primate species are frequently required,[242] however, if investigating the mechanism of a human enzyme, one would first be inclined to express such a target recombinantly within *E. coli*,[243] or if unsuccessful *S. cerevisiae*,[244] then insect cells,[245] and only if explicitly required, mammalian cell lines.[246,247] Bacterial enzymes and proteins are commonly expressed within *E. coli* cells, as their frequent lack of requirement for *post*-translational modifications renders them more easily expressed.[240] Complex proteins, however, do often require co-expression with chaperones or regulatory enzymes that are absent in *E. coli,* to enable production in a fully folded functional, or soluble, form.[248]

### 3.1.2 The Heterologous Expression of Natural Products

The study of NP biosynthetic pathways outside of their cognate host organisms has long presented a significant challenge in the field of biosynthesis. Polyketides and other NPs are often highly complex molecules that are extremely difficult, if not impossible, to access *via* synthetic methodology.[108] Due to their therapeutic potential, large scale fermentation of NP producers has become the method of choice for the production of numerous NP based active pharmaceutical ingredients. For example, *S. erythraea* has been used to produce erythromycin in industrial quantities for as long as this antibiotic has been in clinical use.[249] This parent strain has been 'improved' through multiple cycles of optimisation over many years.[250] Despite recent successes in producing this antibiotic in *E. coli*.,[251,252] *S. erythraea* remains the preferred route to industrial scale erythromycin production. In general, complex NP producers, largely the actinomycetes and related families, are difficult to culture under standard laboratory conditions.[248] We often have a limited understanding of the correct culturing conditions required to both grow and facilitate NP production from host strains.[253] Even in instances where laboratory scale culture has been proven viable, producers are often extremely slow growing and, in most cases, the molecular tools required for genetic manipulation are poorly developed, if at all.[160,248] In addition, the advent of functional genomics, previously discussed in its relevance to the discovery of novel BGCs, enzymes and products, has allowed us to identify multiple BGCs that are not expressed under laboratory conditions.[254] Heterologous expression of BGCs in more tractable model organisms is, therefore, generally preferable to working with the native producers themselves.

The selection of a suitable host organism for heterologous NP pathway expression is challenging. In order to produce a functional PKS or NRPS, the host organism must be capable of

expressing and folding large polypeptide chains encompassing multiple enzyme active sites. In addition, they must be capable of supporting *post*-translational modification of these proteins, through the introduction of the phosphopantetheinyl arm necessary for ACP function.[108] To facilitate the heterologous production of PKS NPs a number of *Streptomyces* species have been investigated, modified and developed as potential surrogate hosts.[255–257] *Streptomyces* are the most well studied of the actinomycetes and hence tend to contain much of the biosynthetic machinery necessary for the correct production of polyketides, as well being a physiologically closer match to the native producer. Early attempts to reproduce PKS function in heterologous hosts centred on *S. coelicolor* and *S. lividans*.[258,259] Many of the current genetic techniques used to facilitate the study of BGCs within *Streptomyces* species were developed using these hosts.[260] Recent studies, however, have sought to use alternative strains to overcome some of the difficulties presented by *S. coelicolor* and *S. lividans*. These alternative *Streptomyces* strains have been selected on the basis of their rapid growth rate and genetic tractability.[257] Many also possess abbreviated genomes,[261] meaning that fewer genetic modifications are needed to generate variant host strains with minimal endogenous biosynthetic activities.[256,262]

### 3.1.3 *E. coli* as a Heterologous Host for the Production of PKS Enzymes and Products

Despite significant progress in strain optimisation, heterologous *Streptomyces* expression hosts are still some way from being the ideal chassis microorganism for PKS expression. Their comparatively slow growth rate, along with the modest toolbox of genetic parts and techniques available for their manipulation and their large genomes rich in BGCs, all represent significant obstacles that still need to be overcome.[263] *E. coli*, by contrast, is in principle an ideal heterologous host. This bacterium is fast growing, is highly amenable to genetic manipulation using well established tools and possesses a well characterised metabolome, allowing for standardised characterisation of NPs produced from this host.[263,264] *E. coli* is, however, not without its limitations. It is physiologically distant from the actinomycetes, and lacks the *post*-translational machinery required to catalyse phosphopantetheinyl transfer.

In an effort to circumvent these issues, Pfeifer *et al.* generated a genetically engineered strain of *E. coli* termed BAP1.[108] The feasibility of using this strain to produce intact PKSs was demonstrated through the expression of a functional DEBS pathway in this host, which in turn delivered modest quantities of 6-dEB, with yields subsequently improved, *via* strain engineering, to levels comparative to heterologous *Streptomyces* hosts.[265] In order to achieve this,

Pfeifer *et al*. increased the quantities of the DEBS substrate, methylmalonyl-CoA, present in the cell by upregulation of the propionyl-CoA synthetase, deleting a propionate catabolism gene cluster and co-expressing a propionyl-CoA carboxylase. Crucially, the team also inserted the gene for a promiscuous phosphopantetheinyl transferase enzyme from *Bacillus subtilis* into the *E. coli*, under the control of the T7 inducer.

The generation of *E. coli* BAP1 has facilitated many of the most significant advances in the understanding of PKS (bio)chemistry, as well as proving a key tool in efforts to re-engineer PKS enzymology. Without exception, the structural characterisation of intact PKS modules from the erythromycin and pikromycin pathways has relied on the use of *E. coli* BAP1 to generate functional protein.[141–143] Kalkreuter *et al*. used *E. coli* BAP1 to produce pikromycin modules capable of installing sequential non-natural extender units[162] and Zhang *et al*. used it in the production of erythromycin analogues.[266] BAP1 cells have also been used to investigate and improve the function and interactions of specific domains within PKSs[155,267] and to investigate substrate selectivity *in vivo*.[268] In addition to the well-studied pikromycin and erythromycin PKSs, BAP1 has also been deployed successfully in investigations of the minimal fungal polyketide synthase LovB[269] hybrid fungal polyketide synthase modules[270] and type I iPKS modules.[271] This strain has also been used to express functional NRPS modules, both alone and as components of hybrid systems.[272] An overwhelming body of evidence suggests that, currently, *E. coli* BAP1 or analogous strains,[170] represent the optimum host system for the recombinant production of PKS polypeptides.

### 3.1.4 Additional Challenges in the Heterologous Expression of mPKSs

Despite its advantages, *E. coli* BAP1 is still some way from being a universal chassis for the heterologous expression of mPKS modules.[264] Notwithstanding the success a small number of labs and their alumni have had interrogating PKS modules from the erythromycin and pikromycin pathways,[108,141,148] there remains a dearth of information regarding PKS modules from non-model systems. Non-model gene clusters are often identified in, and isolated from, a more diverse set of microorganisms, which themselves are often isolated from diverse and evermore hostile environments.[57,58] Many of these strains are extremely slow growing or impossible to culture, and even when this is possible, the isolation of good quality genomic DNA from *Streptomyces* and related species can be challenging.[223] Even when successful genomic DNA purification is achieved, it can still be extremely challenging to work with. Modular PKSs are by

definition repetitive, especially if multiple modules are encoded on a single gene. This presents challenges when attempting to use standard molecular cloning techniques such as PCR or homologous recombination, due to the mismatch of primers or homologous regions of sequence. These issues can be further complicated by very high (or very low) GC content within the genome. For example, the genes *abyB1*, *abyB2* and *abyB3* from the abyssomicin biosynthetic pathway have an average GC content of 76 %. Not only does this create significant repetition within the genes, but it also leads to a necessary codon usage very different to that of *E. coli*.

Codon bias is a well-documented phenomenon, where bacteria use a sub-set of codons far more frequently than others to encode specific amino acids.[273] Bacterial species can display remarkably different biases towards certain codons, which is reflected by the availability of the cognate tRNA within the cell. This bias and availability of tRNAs was identified as the most important factor in determining successful heterologous expression.[274] One potential route towards solving this issue is to take an approach as outlined by Maschio *et al*.[223] This involves commercially synthesising genetic material, codon optimised for the target producer, encoding each PKS polypeptide in as a series of DNA fragments, which are then assembled into an intact coding sequence, to be subsequently inserted into an expression vector and transferred to *E. coli* BAP1. This approach could expedite the recombinant expression of intact polyketide synthase modules, work which will be outlined in this results chapter.

## 3.2 Methods

All general molecular biology methods were performed as described in Chapter 2, with any deviation from these methods highlighted in the text, such as the use of an alternative DNA polymerase. Experimental methods specific to the study outlined in this chapter are outlined below.

### 3.2.1 Colony PCR

Colony PCR screening was performed using a modified version of the general PCR protocol described in Chapter 2, Section 2.1. Individual colonies were picked from the agar plates using sterile pipette tips. Tips were subsequently dipped into 5 µL of $ddH_2O$, before being submerged in 10 mL of LB broth supplemented with Carb. The 5 µL $ddH_2O$ was subsequently used as a template for PCR screening reactions. Screening PCRs were performed using the protocol outlined in 2.1, employing GoTaq Green master mix (Promega). PCR conditions were amended to include an initial 5 minute denaturation step at 95 °C, to ensure comprehensive lysis of the bacterial cells present in the template mixture. Resultant PCR products were analysed by agarose gel electrophoresis as described previously.

### 3.2.2 Small Scale Protein Expression Tests

Agar plates and growth media used for the propagation of *E.coli* transformants harbouring pOPINF derived plasmids (developed by the Oxford Protein Production Facility, UK, specifically for use with the InFusion cloning kit used[275]) were supplemented with 100 µg/mL Carb, whilst plates and media used for the growth of cells transformed with pET28a derived plasmids were supplemented with 50 µg/mL Kan. Cultures were inoculated and grown as described in Section 2.3.1. Further details of the conditions explored during expression testing are summarised in Section 3.4.1. Following overnight growth, cells were harvested by centrifugation at 3,400 x g for 20 minutes. Harvested cells were subsequently re-suspended in 2 mL of His Load buffer. 1.5 mL of this suspension was lysed by sonication on ice. Suspensions were then clarified by centrifugation at 19,000 x g for 20 minutes at 5 °C. Recovered supernatants and pellets were separated and analysed by SDS-PAGE. These samples were prepared as follows: for

supernatants, 2 μL samples were mixed with 5 μL of SDS-PAGE loading buffer before denaturation at 95 °C for 5 minutes; for pellets, a scraping of the cell pellet was taken and resuspended in 100 μL of 8 M urea, with 1 μL of this sample then added to 9 μL of SDS-PAGE loading buffer before denaturation at 95 °C for 10 minutes. All samples were loaded and electrophoresed as described previously.

### 3.2.3 Tryptic Digest MALDI Mass Spectrometry of AbyB3

The identity of recombinant AbyB3 was confirmed using tryptic digest MALDI mass spectrometry. All analyses were performed by the University of Bristol Proteomics Facility. Individual SDS-PAGE gel bands with masses corresponding to that of AbyB3, were excised, washed in dd$H_2O$ and summitted for characterisation.

### 3.2.4 Expression and Purification of AbyB3

AbyB3 was cloned as described in Section 3.4. Following transformation using the heat shock method described in Chapter 2, three colonies of *E. coli* BL21 (DE3) BAP1, transformed with AbyB3::pOPINF, were used to inoculate a single 10 mL solution of LB broth, supplemented with 1 % glucose and 100 μg/mL Carb. 1 L of Modified TB (Melford) medium was prepared according to the manufacturer's instructions, including the addition of 5 mL of glycerol. The media should then be precisely autoclaved, also according to the manufacturer's instructions, as any overheating of the glycerol added results in a lack of expression. The overnight culture is then used as a 1 % inoculant for the prepared media in a baffled flask. The culture is then grown at 37 °C, with shaking at 180 – 200 rpm until an $OD_{600\,nm}$ of ~1 is reached. If no significant growth is observed after 4 h the culture should be disposed of. Once OD has been reached, the culture is placed at 20 °C with shaking at 200 rpm for 1 hour. The culture is then inoculated with a final concentration of 200 μM IPTG and left for 18 h.

The cells were harvested and lysed as described in Chapter 2. The purification also proceeded as outlined in Chapter 2, with the following modifications. AbyB3 was purified by nickel affinity chromatography using a HEPES Load Buffer (50 mM HEPES, 250 mM NaCl, 20 mM Imidazole pH 8) and a HEPES Elute Buffer (50 mM HEPES, 250 mM NaCl, 1 M Imidazole pH 8). It was subsequently purified by SEC using HEPES Buffer (50 mM HEPES,

250 mM NaCl, pH 7.5) and a GE Healthcare 16/60 S200 column. AbyB3 was typically used within 3 days of purification, or flash frozen in liquid nitrogen, in the presence of 10 % glycerol, for storage at -70 °C.

### 3.2.5 Native PAGE Analysis

Protein samples to be analysed by blue native PAGE were mixed with 4x NativePAGE sample buffer (ThermoFisher Scientific) on ice. 500 mL Novex NativePAGE Running Buffer and 200 mL Novex NativePAGE Cathode Buffer (both 20x, Invitrogen) were prepared according to the manufacturer's instructions. A Novex Tris-Glycine Native PAGE gel (ThermoFisher Scientific) was prepared by first rinsing and then filling each well of the gel with Cathode Buffer, prior to transfer of the gel into the electrophoresis apparatus. Protein samples were then loaded, and the central reservoir of the apparatus filled with 1x Cathode buffer. The outer reservoir of the gel apparatus was filled with 1x Running Buffer. Electrophoresis was conducted at 150 V for 100 minutes. Gels were rinsed with dd$H_2O$ and de-stained as outlined for SDS-PAGE gels.

### 3.2.6 Negative Stain Electron Microscopy

AbyB3 was purified by SEC using a non-phosphate based buffer, typically 50 mM HEPES, 250 mM NaCl, pH 7.5. A 300-mesh nickel grid, coated with an ultra-thin layer of carbon, ~ 5 nm thick (Electron Microscopy Sciences), was glow-discharged at 15 mA for one minute using a Leica EM ACE600. On a sheet of parafilm, two 50 µL drops of HEPES buffer and one 5 µL drop of 3 % uranyl acetate solution were prepared. The glow-discharged grid was carefully taken by the rim, and a 5 µL drop of protein solution, at a concentration of 10 µg/µL was applied and left for 2 minutes. After this incubation, the remaining protein solution was removed by blotting using a Whatman No.1 Filter paper. Following blotting, grids were placed face down on a drop of HEPES buffer, followed by a second blotting step to remove excess buffer. Grids were then placed face down on the drop of uranyl acetate, for one minute. Excess uranyl acetate was removed by blotting and grids washed again in HEPES buffer before a final blotting step, and were then left to dry completely at room temperature. Grids were imaged using a 200 kV Technai T20 Microscope, equipped with an Eagle 4 x 4k CCD camera, using a magnification of 60,000 x.

## 3.3 Results: *aby*B1-3 Gene Cloning

### 3.3.1 Domain Assignment and Sequence Analysis

The *aby* mPKS genes were originally identified and reported by Gottardi *et al*.[220] This study also reported a bioinformatic based assignment of individual modules and domains within each of the genes *aby*B1, *aby*B2 and *aby*B3. No subsequent studies have been published with focus on the characterisation of these genes or their gene products. In order to interrogate these ORFs further, their sequences were subjected to additional bioinformatic characterisation. Due to their inherent size and complexity, it was decided to focus efforts on *aby*B2 and *aby*B3 at least in the first instance.

Domain assignments were conducted using an online tool developed by Bachmann and Ravel,[276] allowing for the construction of a schematic map ascribing individual amino acid residues to specific domains, or *inter*-domain and *inter*-module linkers (Figure 3.1). It should be noted that this study predates the more recent recommendations of Keatinge-Clay,[169] and proposals around domain assignment based upon evidence of improved *in vitro* function.[170] In addition, it is still unclear whether these new boundary definitions facilitate or hinder the purification and structural study of mPKS modules, which was the primary aim of the project.

The assignment studies conducted revealed that, in general, each domain within individual Aby modules is flanked by linker regions, which are known, based on studies of other PKSs, to be flexible, thus allowing the movement of domains and the passage of the substrates and intermediates between synthase active sites.[115,116] The 20 amino acid linker between the ACP of M5 and the KS of M6 is of particular interest. Given the necessity for the ACP to travel between the KS and AT of M5 and the KS of M6, this short linker is a clear indication of the close spatial arrangement of M5 and M6. Other identified short linking sequences indicate relative constraint between the inactive DH and KR domains in module 5 and the ER and KR domains in module 6. There are sizeable linkers of > 75 amino acids between each of the three domains in AbyB3. Amino acid sequences for AbyB2 and AbyB3 are provided for reference in appendices A2.2 and A2.3.

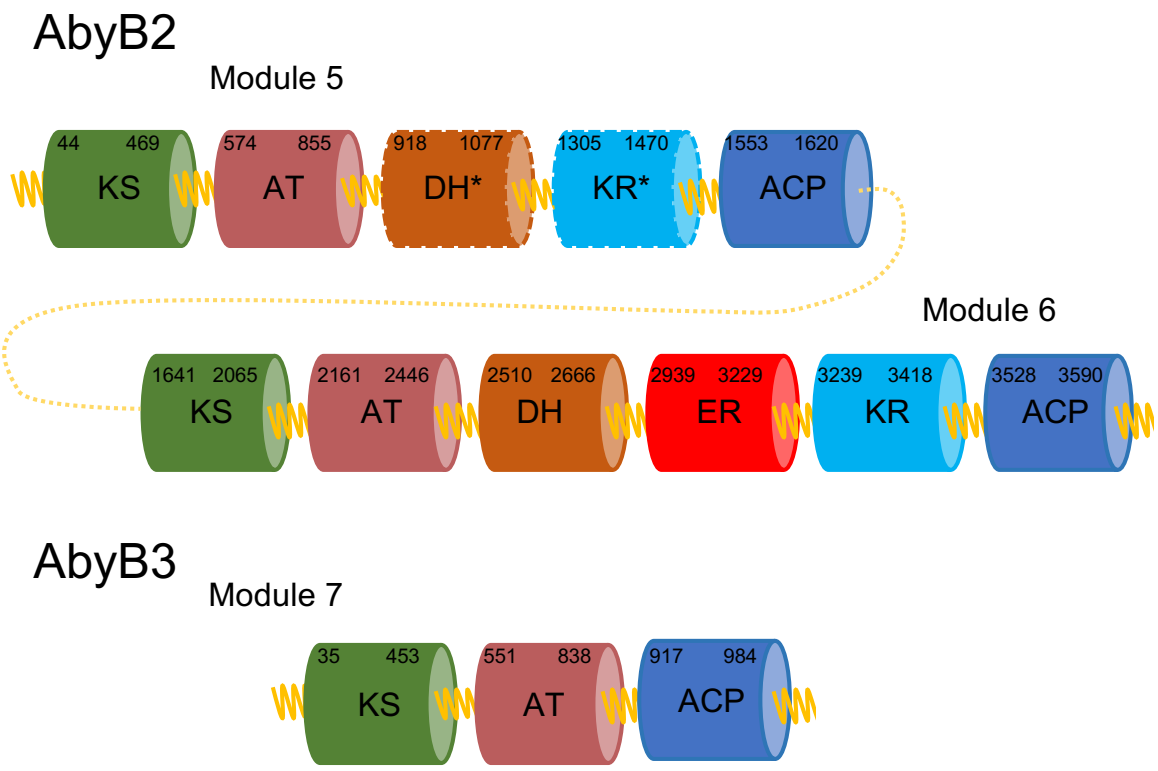



**Figure 3.1 Schematic Diagram Illustrating the Domain Architectures of AbyB2 and AbyB3.** The domain architecture of AbyB2 (3645 amino acids) and AbyB3 (992 amino acids) is shown with amino acid sequence numbering included for clarity. In an alignment consistent with Keatinge-Clay's proposal[169] the ACPs would each be ascribed to the upstream neighbouring module. *Pre*-KS docking domains, *post*-ACP dimerisation helices and *inter*-domain linkers are shown in yellow, with the *inter*-module linker in AbyB2 indicated by a dotted line. Sequence numbering is based on that defined by Bachmann & Ravel's analysis tool.[276] Inactive domains, as identified by Gottardi et al.,[220] are indicated by an asterisk and a surrounding dotted outline.

### 3.3.2 Design and synthesis of DNA fragments

In order to study individual Aby PKS polypeptides, and to facilitate easier downstream engineering approaches, the heterologous expression of the *aby*B2 and *aby*B3 genes was prioritised. The first challenge to overcome was that of the considerable length of the individual genes encoding these proteins. Attempts to clone the entire *aby*B2 gene *via* a conventional PCR amplification approach proved unsuccessful, likely due to the length (11 kb) and high (76%) GC content of this ORF (Appendix A2.1). In order to mitigate this, a strategy was devised involving division of this gene into six separate fragments, each of ~1.8 kb in size. Attempts to

amplify these fragments from *M. maris* genomic DNA proved unsuccessful, with multiple mis-priming events observed. This was a likely consequence of the high GC content and repetitive nature of these sequences.

| Fragment | Nucleotides | Fragment Size (bp) |
|---|---|---|
| AbyB2_Frag1 | 1 - 1844 | 1844 / 1874 |
| AbyB2_Frag2 | 1845 - 3684 | 1840 / 1870 |
| AbyB2_Frag3 | 3685 - 5530 | 1846 / 1876 |
| AbyB2_Frag4 | 5531 - 7373 | 1843 / 1873 |
| AbyB2_Frag5 | 7374 - 9199 | 1826 / 1856 |
| AbyB2_Frag6 | 9200 - 10 938 | 1739 / 1773 |

**Table 3.1 Fragments Designed to Span the Entirety of the abyB2 Gene.** Fragments were synthesised commercially by ThermoFisher GeneArt. The length of the fragments when amplified with the required forward and reverse primers for Type IIs assembly is shown in addition to the fragment size (See Table A1 for primer list).

Following repeated unsuccessful attempts at DNA amplification from *M. maris* genomic DNA, six synthetic codon optimised DNA fragments spanning the entire abyB2 gene were ordered from a commercial supplier (ThermoFisher GeneArt; Table 3.1). Such an approach has only recently become viable due to the rapidly decreasing costs of gene synthesis.[277] This approach presents a number of significant advantages. Firstly, codon optimisation can be performed (Appendix A2.1), prioritising codons most frequently used by *E. coli*, our expression host organism of choice. This can be used to significantly reduce the GC content of a target gene; in the case of *aby*B2, from 76.0% to 56.6%. This both reduces the melting temperature of the DNA, making PCR reactions more tractable (at 76% GC content only 2-step PCR reactions are possible), and reduces the likelihood of mis-priming events.

### 3.3.3 Assembly of the *aby*B2 Gene

Despite the reduction in repetition and GC content, assembly *via* a standard homologous recombination approach using the In-Fusion system (Takara),[278] gave only truncated AbyB2

coding sequences. These were found to be being generated due to the sequence complementarity between AbyB2_Frag1 and AbyB2_Frag6. In an effort to circumvent this problem, a golden-gate based assembly method was employed (Figure 3.2).[279–281] This assembly method uses the ability of Type IIs restriction enzymes, in this case *Bsa*I, to cut DNA at a defined distance away from their recognition site. *Bsa*I does this whilst also generating a four nucleotide 5' overhang. This means that, with careful design, fragments can be created that generate highly specific overhangs, which are complimentary to equivalent terminal sequences in neighbouring DNA fragments sequences. The BsaI recognition site is neatly excised during this process, allowing the formation of a scarless join when incubated with T4 DNA ligase.

To create the necessary recognition sequences and overhangs, the ThermoFisher GeneArt Primer and Construct Design tool was used. This automatically analyses different possible combinations of overhangs, excluding those which present too great a chance of mis-priming. It also designs overhangs for a recipient vector, pType-IIs, specifically designed for use in these reactions and provided with the Type IIs assembly kit sold by ThermoFisher GeneArt. The pType-IIs vector contains ampicillin and chloramphenicol resistance genes, as well as restriction sites for the Type IIs enzymes *Bsa*I, *Bbs*I and *Aar*I. During this study, a one pot reaction, involving a master-mix of buffer, T4 DNA ligase, *Bsa*I and all six *aby* DNA fragments, PCR amplified to contain the necessary recognition and overlap sequences, was performed at 37 $^{o}$C for 1 hour, followed by 5 minutes at 55 $^{o}$C. This reaction mix was then used to transform chemically competent *E. coli* DH5α Stellar cells (Takara).

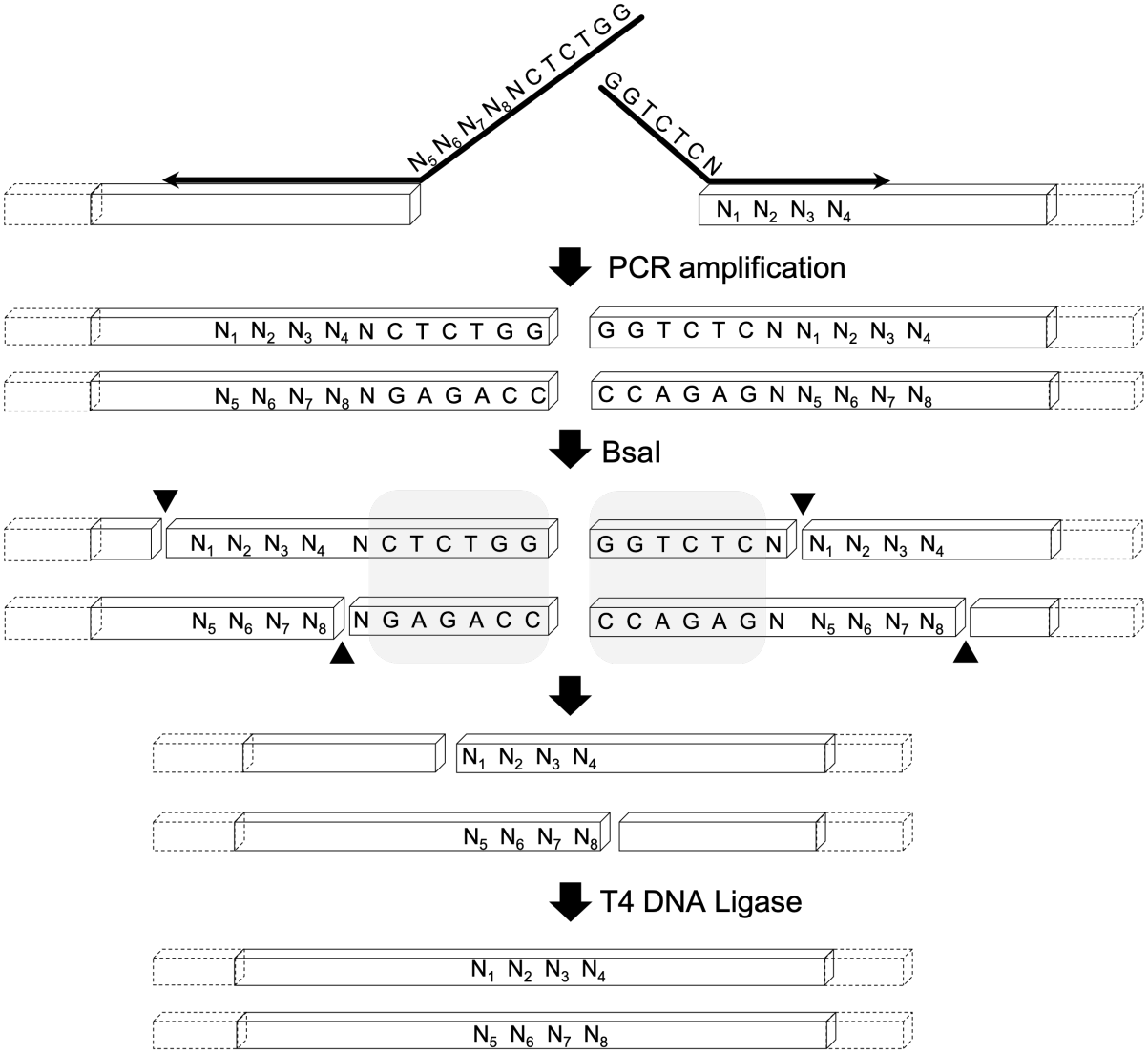



**Figure 3.2 Schematic Representation of the Assembly of the abyB2 Gene.** Each synthesised *aby*B2 fragment was amplified by PCR using the requisite primers to generate a *Bsa*I restriction site (Table A1) and a unique homology region, compatible with upstream and downstream fragments. Fragments and the pType-IIs vector are digested to generate the requisite overhangs, following which all six fragments are assembled in the correct order in the same one-pot reaction. *Bsa*I restriction sites are highlighted in grey.

A number of minor modifications were made to recommended protocols to ensure correct assembly of the *aby*B2 gene at a reasonable cost. Although the assembly of > 4 fragments is recommended to be undertaken using discrete shuttle vectors containing each fragment, with assembled genes recommended to be no more than 10 kb in length, we opted to disregard these recommendations in an effort to streamline the assembly process. Due to the expected low

assembly rates and reduced fidelity, a total of 24 *E. coli* transformants were recovered from agar plates and screened by colony PCR (Figure 3.3). This enabled the identification of the presence of a single 1.5 kb section of the *aby*B2 gene, spanning the AbyB2_Frag2 and AbyB2_Frag3 join, thus confirming the presence of at least two fragments assembled in the correct order. Plasmid DNA from the five most promising candidate transformants (5, 6, 9, 13 and 18; Figure 3.3) was miniprepped and subjected to further analysis.

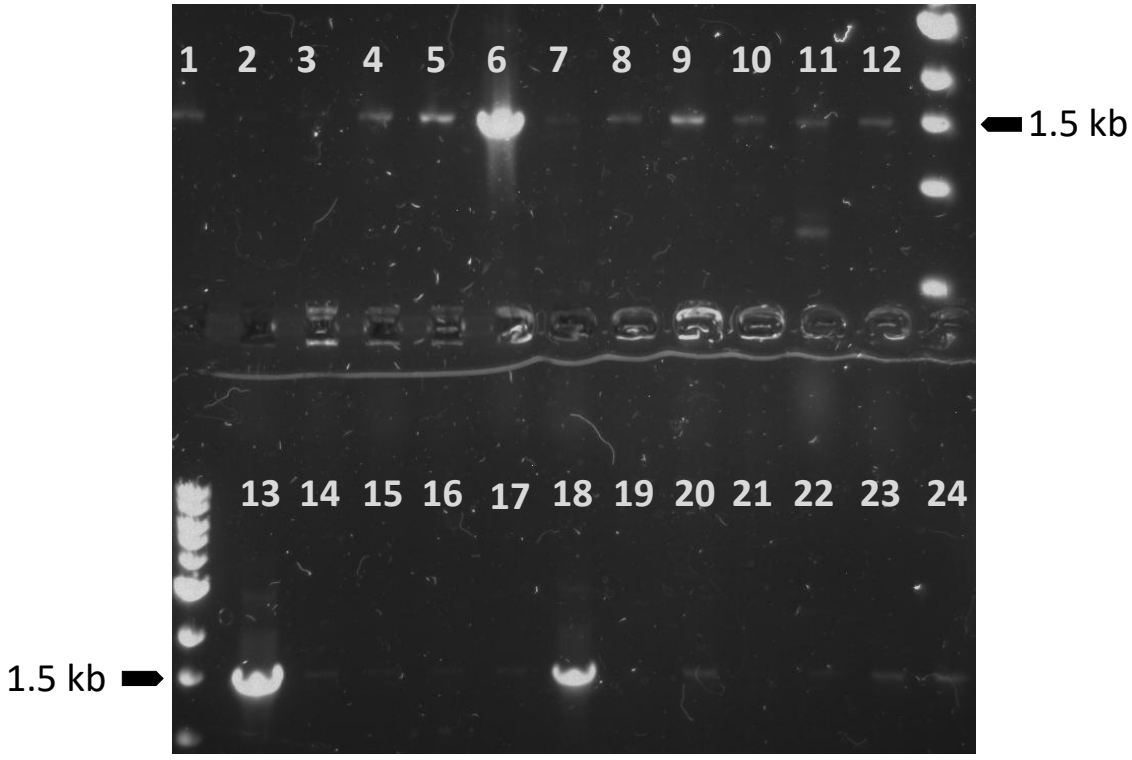



**Figure 3.3 Colony PCR Screening of Fragment Assembly and Insertion into pType-IIs.** Twenty four colonies were picked and the presence of a 1517 bp fragment, spanning the join between AbyB2_Frag2 and AbyB2_Frag3, was confirmed by colony PCR.

Purified plasmids were subsequently used as the template for further PCR analysis. Amplification reaction were performed to confirm the presence of 1.5 kb sections of *aby*B2, spanning AbyB2_Frag1 and AbyB2_Frag2, AbyB2_Frag3 and AbyB2_Frag4, and AbyB2_Frag5 and AbyB2_Frag6 (Figure 3.4). This enabled confirmation of the presence of each fragment within the gene, as well the correct order between AbyB2_Frag1 – AbyB2_Frag4 and AbyB2_Frag5 – AbyB2_Frag6. A final, full length, PCR reaction (Figure 3.5) then enabled confirmation of

the connection of AbyB2_Frag1 and AbyB2_Frag6 to pType-IIs, and hence the fidelity of the order for all fragments within the assembled ORF.

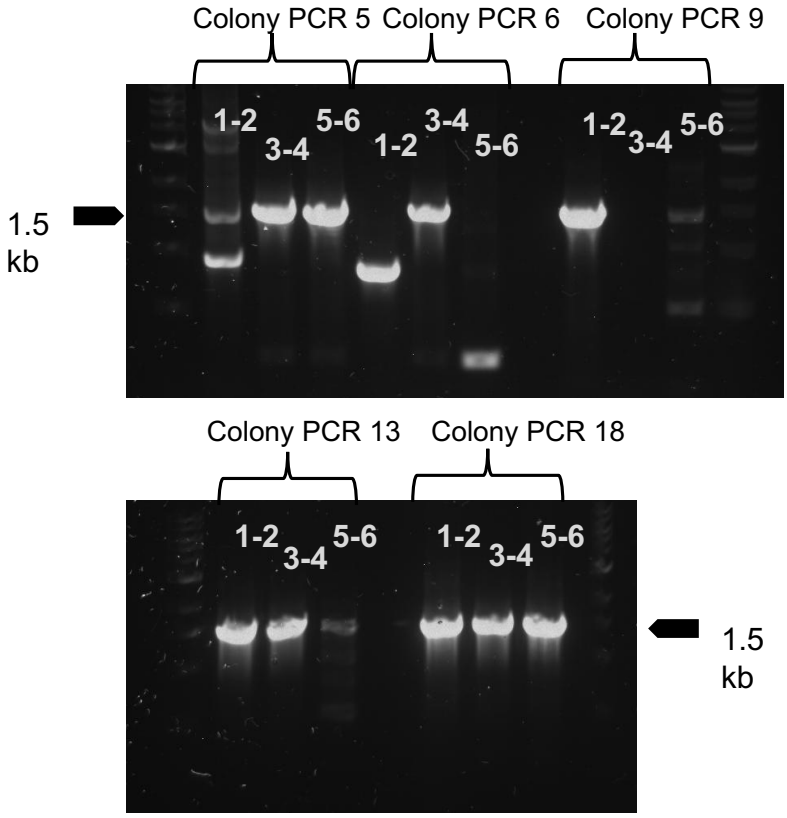



**Figure 3.4 Verification of the Presence of abyB2 DNA Fragments by PCR.** Plasmids were isolated from the five candidate colonies previously shown to contain AbyB2_Frag2 and AbyB2_Frag3. The presence of all fragments was determined by PCR amplification of ~1.5 kb fragments spanning the join between AbyB2_Frag1 and AbyB2_Frag2 (1-2), AbyB2_Frag3 and AbyB2_Frag4 (3-4), and AbyB2_Frag5 and AbyB2_Frag6 (5-6).

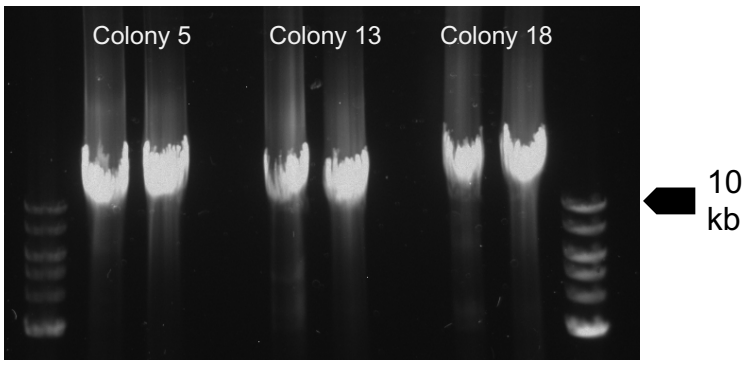



**Figure 3.5 PCR Amplification of the Intact *aby*B2 Synthetic Gene.** *aby*B2 was amplified in duplicate from the plasmids obtained from colonies 5, 13 and 18 by PCR, using LongAmp Taq DNA polymerase (NEB). Amplified fragments were purified and inserted into pOPINF or pET28 by homologous recombination.

Following assembly, the intact synthetic *aby*B2 gene was inserted into either the expression vector pOPINF or pET28A (Table A1.2), with successful insertion confirmed by colony PCR from plasmid transformants recovered from agar plates. Plasmids originating from colonies 13 and 18 were submitted for DNA sequencing. Sequencing primers were designed in both forward and reverse directions, complementing central DNA sequences within each *aby*B2 gene fragment. In addition, the pType-IIs plasmids containing genes 13 and 18 were sent for sequencing in the forward direction (using all designed forward primers and T7 forward, Table A1), whilst equivalent pOPINF constructs were sent for sequencing in the reverse direction (using all designed reverse primers and T7 Term). Sequencing results revealed all six fragments were assembled in the correct order. The *aby*B2::pOPINF plasmid derived from colony 18 was determined to be >99.9% sequence correct, with the remaining 0.1 % of the gene sequence equating to regions where there was poor sequencing coverage.

### 3.3.4 Cloning *aby*B3

The *aby*B3 gene, at 2979 bp long, is much shorter than *aby*B2, which allowed the synthesis of a single codon optimised coding sequence (ThermoFisher GeneArt). In addition, the gene could be successfully amplified from genomic DNA, despite its high (75%) GC content (Figure 3.6). Genomic and codon optimised *aby*B3 was inserted into pOPINF by homologous recombination in preparation for expression trials.

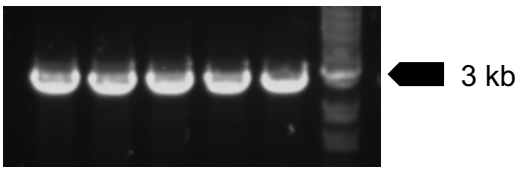



**Figure 3.6 PCR amplification of abyB3 from genomic DNA.** A temperature gradient during the annealing step between 68 – 72 °C (left to right on the gel) ensured effective sequence amplification.

## 3.4 Results: Heterologous Expression of AbyB2 and AbyB3

### 3.4.1 Protein Expression Trials

The plasmids *aby*B2::pOPINF, *aby*B2::pET28a, *aby*B3::pOPINF and *aby*B3::pET28A were transformed into a variety of *E. coli* cell lines for expression testing (Table 3.2). Once the required cell was transformed, expression of the target polypeptide was assessed under a variety of expression conditions, including the in the presence of varying concentrations of IPTG, at different growth temperatures, and by varying the length of time that cells were left to grow *post*-induction. Initially, *E. coli* BL21 (DE3) cells were trialled for expression, as these have previously been used successfully for the production of inactive mPKS polypeptides in the erythromycin and pikromycin systems, without the insertion of a phosphopantetheinyl arm. The majority of expression trials for AbyB3 in *E. coli* BL21 (DE3) cells and *E. coli* BL21 (DE3) pLysS cells, were undertaken by Dr. Alice Parnell, although these unfortunately failed to generate recombinant protein. AbyB2 expression trials in *E. coli* BL21 (DE3) cells and pLysS cells were undertaken by myself, as detailed in Table 2, as were complimentary AbyB2 and AbyB3 expression trials using BAP1 and Vmax cells. These are a commercially sourced strain of *Vibrio natriegens* (SGI-DNA), purported to be a superior expression host to *E. coli*, delivering higher yields of overexpressed protein per litre of culture and with a faster doubling time.[282,283] Initially, these cells were only available in electrocompetent form; successful transformation proved all but impossible, despite fastidious adherence to the manufacturer's protocol. The later release of chemically competent cells did little to improve transformation efficiency, prohibiting a full characterisation of expression conditions to be undertaken using this cell line. *aby*B2::pET28a and *aby*B3::pOPINF were the only plasmids that could be successfully transformed into this cell line, though in both cases there was no evidence of recombinant protein expression following growth in the presence of the chemical inducer IPTG.

| Protein | Plasmid | Host cell | Media | Temperature post-induction (°C) | Final IPTG concentration (μM) | Time post-induction | Notes |
|---|---|---|---|---|---|---|---|
| AbyB2 | pOPINF | BL21 (DE3) | LB | 20 | 500 | 4h, O/N | With and without 10 min ice shock pre-induction |
| AbyB2 | pOPINF | BL21 (DE3) | TBAI | 37 | - | 4h, O/N | With and without 10 min ice shock pre-induction |
| AbyB2 | pOPINF | pLysS | LB | 20 | 500 | 4h, O/N | |
| AbyB2 | pOPINF | pLysS | TBAI | 37 | - | 4h, O/N | |
| AbyB2 | pET28 | BL21 (DE3) | LB | 20 | 500 | 4h, O/N | |
| AbyB2 | pET28 | pLysS | LB | 20 | 500 | 4h, O/N | |
| AbyB2 | pET28 | Vmax | VM | 30 | 1000 | 1h, 4h, O/N | |
| AbyB2 | pET28 | Vmax | LBv2 | 30 | 1000 | 1h, 4h, O/N | |
| AbyB2 | pOPINF | BL21 (DE3) | TB | 20 | 200 | 4h, O/N | 1h at 20°C pre-induction |
| AbyB2 | pET28 | BL21 (DE3) | TB | 20 | 200 | 4h, O/N | 1h at 20°C pre-induction |
| AbyB2 | pOPINF | BAP1 | TB | 20 | 200 | 4h, O/N | 1h at 20°C pre-induction |
| AbyB2 | pET28 | BAP1 | TB | 20 | 200 | 4h, O/N | 1h at 20°C pre-induction |
| AbyB2 | pOPINF | pLysS | TB | 20 | 200 | O/N | Co-transformed with *Sfp* plasmid |
| AbyB2 | pOPINF | pLysS | TBAI | 20 | - | O/N | Co-transformed with *Sfp* plasmid |
| AbyB2 | pOPINF | Arctic Express | TB | 12 | 200 | 24 h | Co-transformed with *Sfp* plasmid |
| AbyB2 | pOPINF | Arctic Express | TBAI | 12 | - | 24 h | Co-transformed with *Sfp* plasmid |
| AbyB3 | pOPINF | Vmax | VM | 30 | 1000 | 1h, 4h, O/N | |
| AbyB3 | pOPINF | Vmax | LBv2 | 30 | 1000 | 1h, 4h, O/N | |
| AbyB3 | pOPINF | BL21 (DE3) | TB | 20 | 200 | 4h, O/N | 1h at 20°C pre-induction |
| AbyB3 | pOPINF | BAP1 | TB | 20 | 200 | 4h, O/N | 1h at 20°C pre-induction |

**Table 3.2 Expression Conditions Tested for AbyB2 and AbyB3.** The only conditions observed to give successful protein expression were those that employed *E. coli* BAP1 cells. Additional unsuccessful expression testing of AbyB3 was carried out by Dr. Alice Parnell. Substantial issues transforming Vmax

cells prevented full characterisation. BAP1, BL21 (DE3) BAP1 cells[108] (a gift from Mr. Christian Hobson, University of Warwick); LB, Luria-Bertani broth (Thermofisher); LBv2, LB broth with V2 salts added; O/N, overnight; pLysS, BL21 (DE3) pLysS cells (Agilent); TB, Terrific Broth (Melford); TBAI, Auto-Induction Terrific Broth (Formedium); VM, Vmax Enriched Growth Medium (SGI-DNA); Vmax, *Vibrio natriegens* (SGI-DNA).

Successful expression of both genomic and codon optimised AbyB3 was observed in *E. coli* BL21 (DE3) BAP1 cells (a gift from Mr. Christian Hobson, Warwick; Figure 3.7). This cell line originates from the parental strain lineage generated by Khosla and co-workers.[108] Expression was observed in these cells but interestingly not from standard BL21 (DE3) cells under the same conditions. This leads to the intriguing possibility that, in this system, folding and/or stability of the multidomain complex might somehow be dependent on the insertion of a phosphopantetheine arm and the upregulation of cellular levels of propionyl-CoA, the precursor for the methylmalonyl-CoA extender unit incorporated by AbyB3. Successful expression of AbyB2 was not observed under the same conditions (Figure 3.7). Expression of the correct AbyB3 construct was confirmed by tryptic digest MS, performed by the Bristol University Proteomics Facility. Further investigations into the expression of AbyB2 in the presence of the *B. subtilis* phosphopantetheinyl transferase *Sfp* did not yield positive results, even at low temperatures in Arctic Express cells.

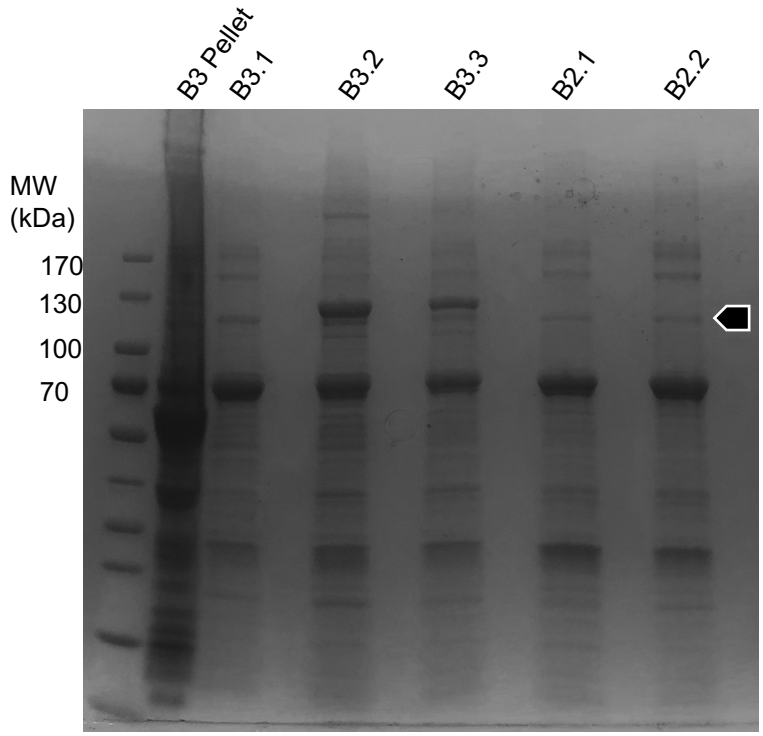



**Figure 3.7 Expression Trials of AbyB2 and AbyB3 in E. coli BAP1 Cells**. Overexpression of AbyB3 is observed for B3.2 and B3.3. The expected position of AbyB3 on the SDS-PAGE gel based on amino acid composition is indicated by an arrow. Overexpression bands in B3.2 and B3.3 were excised and verified to be AbyB3 by tryptic digest and mass spectrometry. B3 Pellet, insoluble pellet produced by B3.2; B3.1, genomic abyB3 DNA inserted into pOPINF; B3.2, genomic abyB3 gene (separate PCR amplification) inserted into pOPINF; B3.3, codon optimised synthetic gene of abyB3 inserted into pOPINF.

### 3.4.2 Large-Scale AbyB3 Expression and Purification

Due to the increased overexpression levels observed for the *aby*B3::pOPINF construct containing the authentic *M. maris* sequence in *E. coli* BAP1 cells, this plasmid and strain combination were taken forward for large scale expression studies. When following the protocol outlined in Section 3.2.4, high level recombinant AbyB3 expression could be readily achieved. This was dependent on following the outlined expression protocol explicitly. Small deviations from this protocol e.g. not growing the BAP1 cells at 20 ºC for 1 h prior to induction, or autoclaving the TB media for an excessive period of time, resulted in a complete loss of protein

expression. In addition, significant variation was observed in expression levels from different colonies recovered from transformation reactions. Initial slow growth at 37 ºC proved to be a reliable indicator that overexpression of AbyB3 would not be observed. Two growths from different colonies picked from the same plate of transformants frequently yielded radically different levels of target protein.

Once a formalised overexpression protocol for AbyB3 had been finalised, purification of this protein was performed using a combination of nickel affinity chromatography and SEC. This yielded recombinant AbyB3 protein of > 95 % purity (Figure 3.8). AbyB3 was observed to resolve at a molecular weight slightly higher than that predicted based in its amino acid composition by SDS-PAGE, indicative of the presence of some regions of intrinsic disorder.

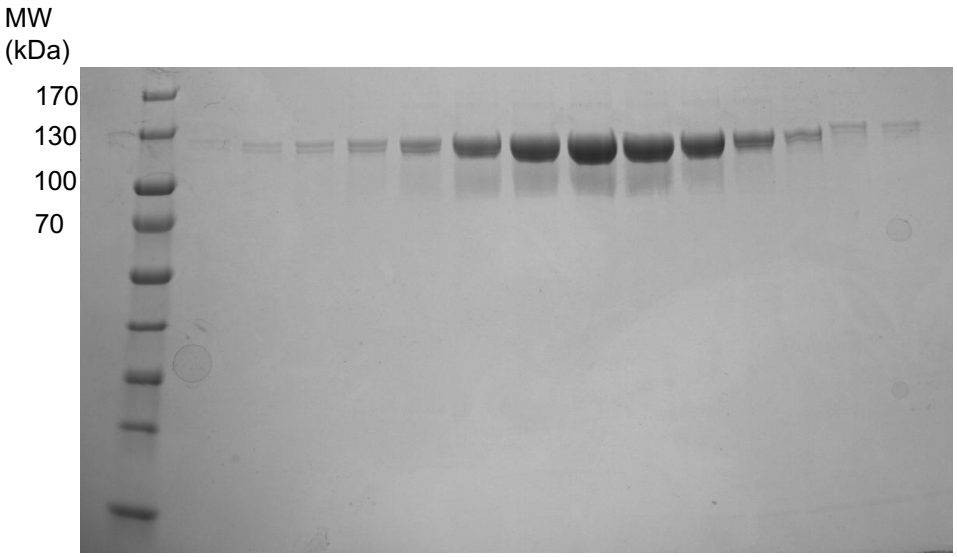



**Figure 3.8 SDS-PAGE gel of purified AbyB3.** Samples from across an AbyB3 elution peak recovered during SEC purification are shown. Samples were run on a 4-20 % SDS-PAGE gel (NuSep), with fractions loaded in the order of elution from left to right.

### 3.4.3 Preliminary Analysis of AbyB3

Given the well-reported potential for mPKS modules to dimerise *via* their KS domains,[113,114] the oligomeric state of AbyB3 in solution was investigated by native PAGE (Figure 3.9). This revealed an unexpected mixture of dimer and tetramer. No tetrameric mPKS species have been reported to date, although it is possible that those that have been successfully expressed and purified have not been subjected to hydrodynamic analysis. The tetrameric form of AbyB3

appeared in all samples of protein analysed following SEC, indicating that there was a dynamic equilibrium between the species, and initial investigations did not reveal large changes to the ratio of dimer to tetramer upon concentration of the sample.

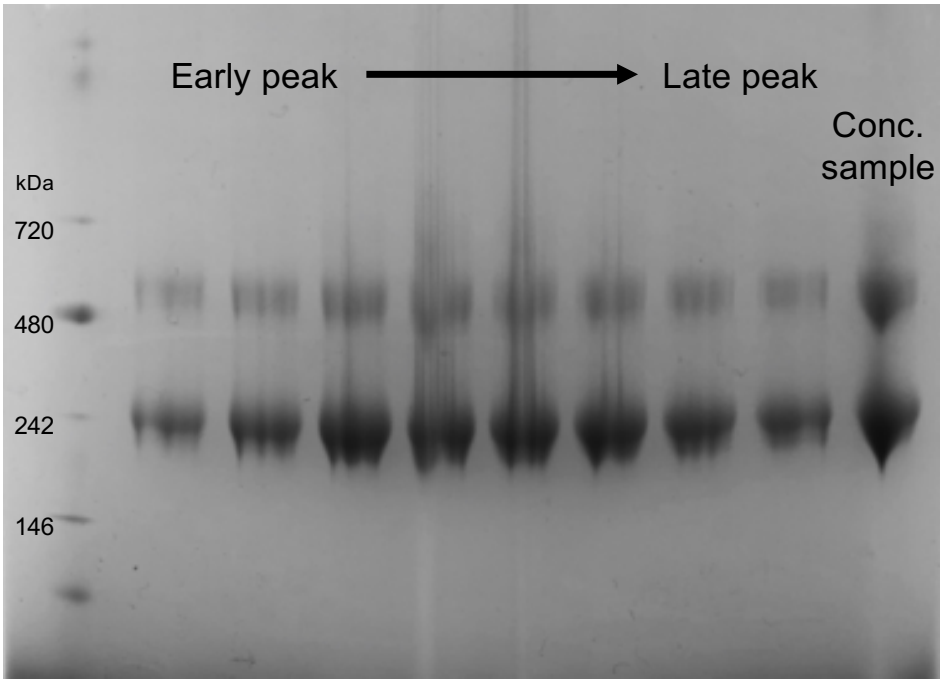



**Figure 3.9 Blue Native PAGE Gel of Purified AbyB3.** Purified AbyB3 was resolved on a native PAGE gel immediately following SEC. Bands are observed that correspond approximately to the expected size of a dimer (213 kDa) and tetramer (426 kDa). Fractions from across the peak were used for analysis to investigate the possibility of different sized populations of AbyB3 being eluted at different times. A concentrated sample was also loaded on the gel to determine the effect of AbyB3 concentration on its oligomeric state.

Further analysis of purified AbyB3 was conducted *via* negative stain electron microscopy (NSEM), and revealed a sample containing what appears to be a mixture of dimeric and tetrameric particles (Figure 3.10). These particles initially look to be pure and homogenous, and exhibit a slightly round appearance, with, in some cases an observable central cavity. These initial results are consistent with those reported by Dutta *et al*.,[141] who similarly observe a compact rounded object, as opposed to an architecture more analogous to that observed in porcine fatty acid synthase.[136]

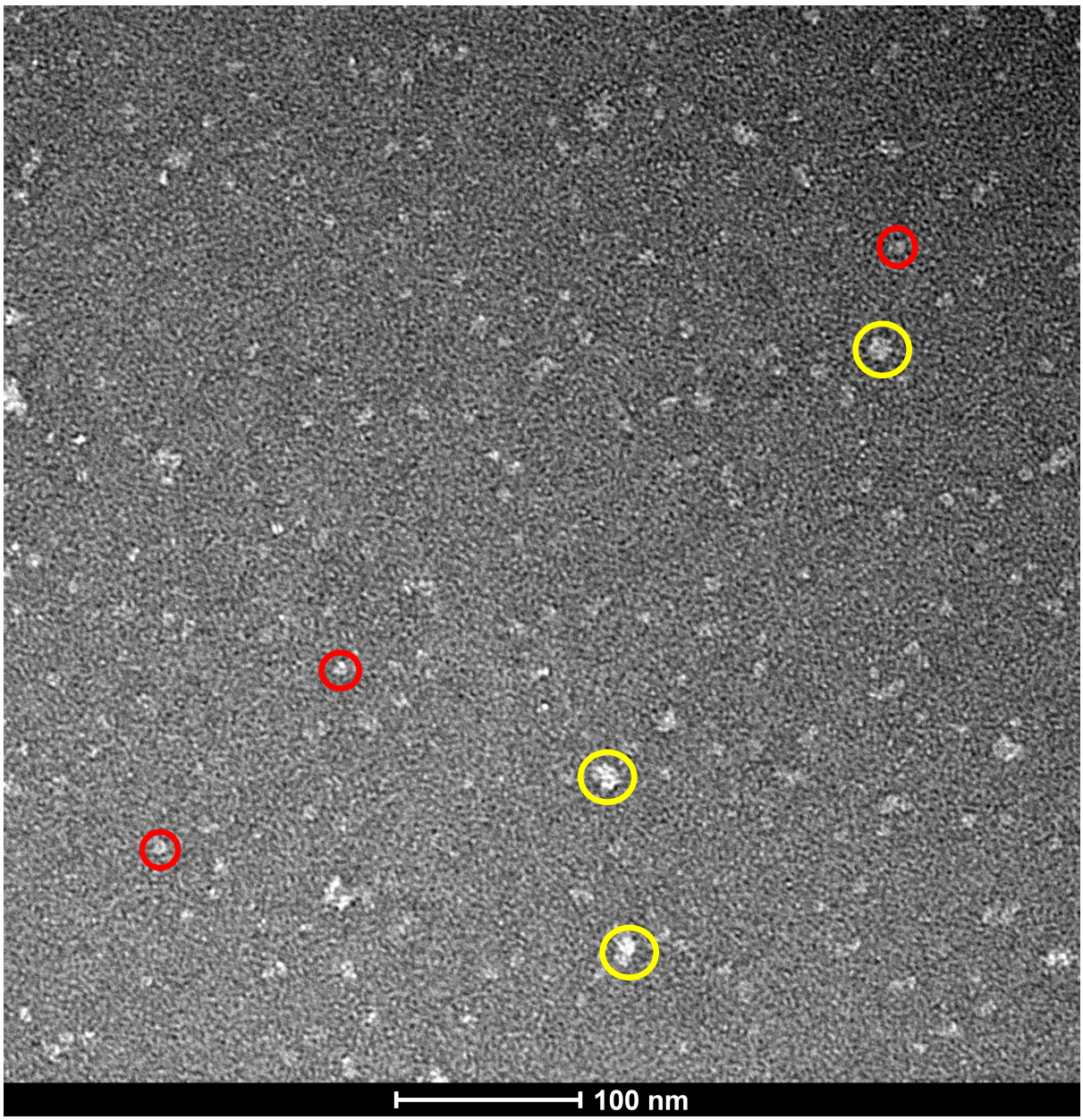



**Figure 3.10 Analysis of Purified Recombinant AbyB3 by Negative Stain Electron Microscopy**. Micrographs were obtained on a Technai T20, with a magnification of 60,000 x. Example dimers (red circles) and tetramers (yellow circles) are highlighted.

## 3.5 Discussion

### 3.5.1 Heterologous Expression of mPKS Modules

The work outlined in this chapter sought to develop a cloning and expression methodology to enable to the high-level production of recombinant multi-domain polypeptides from the abyssomicin C PKS. These studies focused specifically on the proteins AbyB2 and AbyB3 as test cases for the establishment of such a methodology. Due to the size and complexity of these coding sequences it was anticipated that conventional molecular cloning techniques would have limited value, hence a decision to exploit recent innovations in DNA synthesis and assembly. Modern gene synthesis techniques have opened up new avenues from which to approach complex projects involving large genes with repetitive sequences or high GC content. These advances have also translated into significantly reduced synthesis costs,[277] for example, the genes used in this study were synthesised at 14 p per base.

Although significant effort was committed to attempting to clone the *aby*B2 gene from genomic *M. maris* DNA, this ultimately proved intractable and without the option to commercially synthesise this coding sequence the project would have been halted. The codon optimisation performed prior to synthesis is also of considerable value, reducing the protein production burden on the host bacteria. Although recombinant AbyB2 was ultimately not produced during this study, this work is now ideally positioned for future optimisation. The benefit of codon optimisation to increasing the tractability of working with high GC content sequences is also considerable, as is evidenced in this study. It considerably reduces the Tm of any primer designed to work with a target gene, allowing more reliable 3-step PCR reactions to be performed. Nevertheless, despite the synthetic DNA being far easier to work with, fragment assembly by traditional digestion and ligation methods, or even more modern homologous recombination methods proved challenging, at least in the work outlined herein. By contrast, Golden Gate based assembly methods were found to be ideally suited to the demands of this project. By extension, it is apparent that these methods will of major value in facilitating access to a plethora of mPKSs from lesser-known bacterial and fungal pathways.[112,284]

Once expression plasmids for both AbyB2 and AbyB3 had been assembled and verified, a number of expression protocols were trialled in an effort to access recombinant target polypeptides. Unfortunately, the vast majority of the approaches trialled during this project proved unsuccessful. Undoubtedly, the physiological disparities between *E. coli* and *M. maris* will

have impacted on the success, or not, of this work. Despite this, strict adherence to an in-house formulated bespoke expression protocol enabled the production of AbyB3 at in good yield. *E. coli* BAP1 cells not only provide a facile way to express active mPKS modules, and, as evidenced by this study, support the expression of the large, multi-component protein complexes. It is unclear whether, without the added propionyl-CoA carboxylase, cellular levels of methylmalonyl-CoA are increased, and whether this had any impact on AbyB3 expression. If deletion of the propionyl-CoA metabolism cluster, through the insertion of the *Sfp* gene, does facilitate expression, it would be desirable to knockout acetyl-CoA metabolism for the expression of AbyB2. Further expression testing of AbyB2 in *E. coli* Arctic Express and pLysS cells with the addition of *Sfp* through co-transformation, gives good evidence that this alone is not sufficient to enable heterologous expression. The significant variability in AbyB3 expression yields clearly indicates a requirement to operate in a Goldilocks expression window. It highlights the need for rigour when investigating potential expression conditions, as minor changes can have significant effects on the success of the trial. In addition, the variability of the colony picked *post* transformation means that duplicate or triplicate trials should always be run for each change in condition. Together, these data give some confidence that equivalent expression protocols could be readily developed for AbyB2 during future studies.

The successful expression of AbyB3 provided a sound foundation for the production of homogeneous recombinant AbyB3 at scale. Based on these studies a two-step purification protocol was developed for this protein, which delivered > 95 % purity material in sufficient quantities to enable structural characterisation. To date, no standalone minimal PKS module has been successfully expressed and purified for analysis, and only modules from the erythromycin, pikromycin and venemycin pathways have been produced.[108,141,170] This later system, isolated from *S. venezuelae*, is a simple, three-module mPKS which is yet to be subjected to structural or functional characterisation. Notably during this study, it was proven comparatively straightforward to purify AbyB3 to homogeneity, with the inclusion of reducing agents in the SEC buffer used affording greater stability of the protein, thus enabling longer term sample storage. A relatively high yield of AbyB3, and the ability to purify it to homogeneity provides a good platform to proceed to more detailed characterisation.

### 3.5.2 Summary

Initial characterisation of recombinant AbyB3 revealed that the polypeptide adopts an unexpected tetrameric oligomeric state *in vitro*. Although it is unlikely that the tetrameric form of this protein represents the *bona fide* natural conformation, it does make analysis by electron microscopy more viable. For high resolution structural characterisation by cryo-EM, multimeric complexes with more axes of symmetry lead to higher resolution structures, as do larger protein complexes. From the NSEM analysis of AbyB3 reported herein it is evident that there is good potential for further structural characterisation of this polypeptide *via* cryo-EM. Initial inspection of the particles observed by NSEM indicate that the conformation adopted by AbyB3 is consistent with those previously reported by Dutta *et al*. and Whicher *et al*. for PikAIII. With this in mind, a decision was made to initiate a Cryo-EM study of AbyB3, studies which are detailed in the next chapter of this thesis.

# Chapter 4

# Structural Characterisation of AbyB3

## 4.1 Introduction

### 4.1.1 Crystallographic Studies of PKS Didomains

Current structural models of intact mPKS modules are derived from a small number of studies of multi-domain mPKS polypeptides. These studies present conflicting evidence for the conformational arrangement of individual domains within the mPKS module, but each are not without merit. It is the hypothetical interpretations of these data that leads to the reported discrepancies. The first model, that proposed by Khosla and co-workers,[143] and adopted later by Keatinge-Clay and others,[143,170] is primarily based on the crystal structure of the KS-AT didomain from DEBS M5 (Figure 4.1a and b).[113] An analogous crystal structure of the DEBS M3 KS-AT didomain was published soon after by the same group.[114] These high resolution crystal structures enabled a significant step forward in the understanding of both the structure and the function of mPKS domains and modules. For the first time, active sites were identified, and models for the interaction of the ACP with both the AT and KS domains were developed. These advances provided good evidence for the protein-protein interactions postulated to be essential for correct mPKS function.[80] It is unsurprising, then, that Khosla and co-workers used these didomain structures as the basis for the formulation of a molecular model of mPKS architecture, which was supported by SAXS studies of DEBS M3-TE.[143] These analyses was performed using a rigid body fit, as opposed to an ensemble-state calculation, as their data showed that DEBS M3-TE "did not sample a wide range of conformational states". In light of this, it is necessary to examine the crystal structures underpinning these data and the assumptions that were made during analysis. The structure of the DEBS M5 didomain was solved with three dimers in the asymmetric unit (Figure 4.1c), which gives us some idea about the packing

constraints within the crystal that may prevent the didomain from exhibiting a conformation representative of the authentic intact module. When viewed as an entire asymmetric unit we can see significant crystal contacts within the lattice. Whilst this is a common constraint in crystallography, it is well known to restrict the flexibility of target macromolecules.[285] These crystal contacts could, theoretically, prevent the didomain structure adopting a conformation more akin to the arch shaped conformation observed by Dutta *et al*.[141]

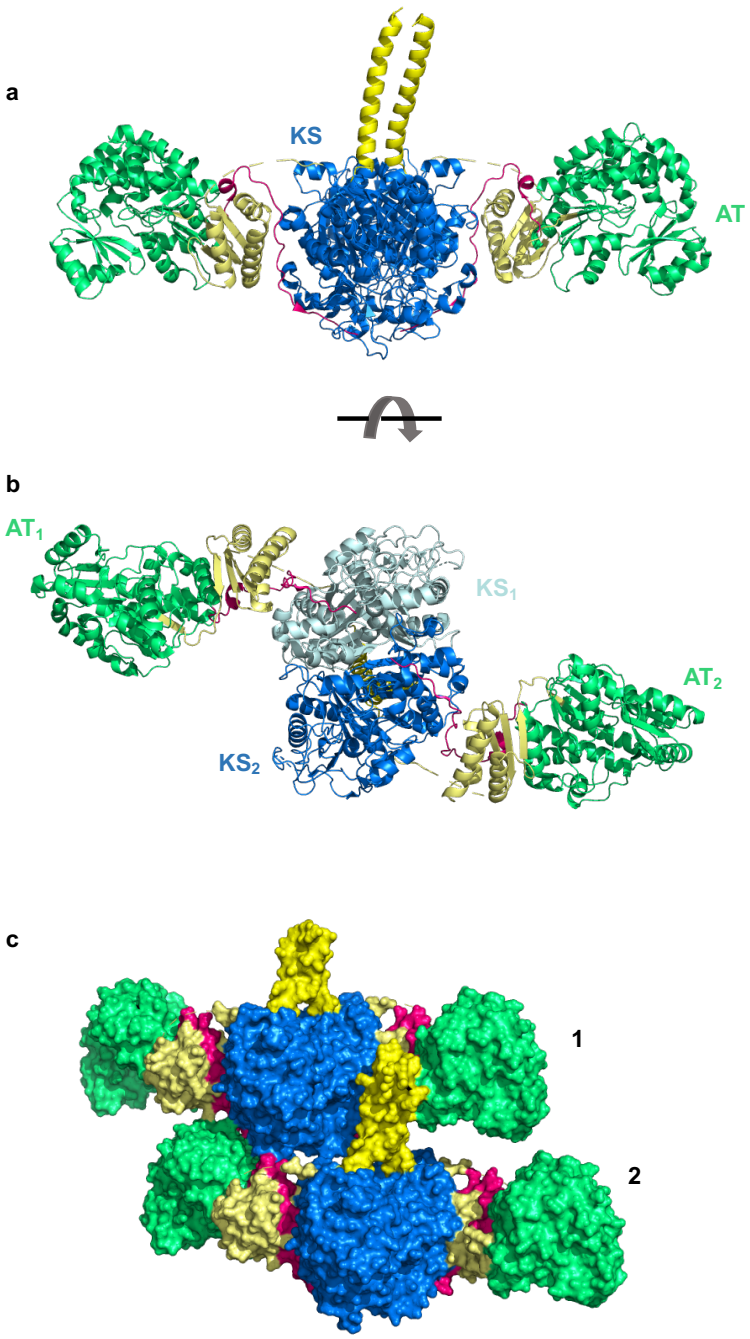



**Figure 4.1** **X-ray Crystal Structure of DEBS M5 KS-AT Didomain.** a) Side-on view of the didomain structure. The pre-KS docking domains are in yellow, the KS dimer is in blue, the KS-AT linker in light yellow, the AT in green and the post-AT linker in pink. b) The same structure rotated 90°. This time the one KS monomer is coloured cyan and the dimerisation interface is visible. c) A solid rendering of two KS dimers in the unit cell, with colouring consistent with part a. Extensive crystal contacts are visible between the KS in dimer 1 and the AT in dimer 2, as well as the pre-KS docking domain in dimer 2 and the groove between the KS and AT in dimer 1.

### 4.1.2 Analysis of SAXS-derived mPKS Module Structures

Edwards *et al*.[143] addressed the issue of didomain confirmation, albeit indirectly, through SAXS modelling of the solution structures of the individual domains contained within DEBS M3-TE (Figure 4.2a). The high resolution crystal structures of the individual domains fit well into their respective SAXS envelopes, though the KS-AT dimer adopts a more planar conformation, as opposed to that imposed in the crystal lattice, with each AT domain angled towards the *pre*-KS docking domain. Whilst this structure gives good evidence that the KS-AT dimer does not adopt a significantly altered conformation in solution, it does hint towards the need to model some flexibility into the linker region between the two domains. In contrast, the SAXS analysis of intact DEBS M3-TE is highly ambiguous. The resolution reported is low, ~40 Å, and there is a high degree of conformational flexibility within the particle. Three distinct conformational arrangements are given for the KR, ACP and TE domains, which are reported to move significantly, relative to each other, in solution (Figure 4.2b). Despite this flexibility, none of the determined structural models appear to convincingly fill the density identified in the SAXS model (Figure 4.2c), with no attempts made to dock high resolution structures of these domains into the experimental SAXS envelope. The KS-AT dimer, however, is consistently modelled as a rigid body, without any flexibility. This is particularly surprising given the flexibility observed in the rest of the complex, the existence of the KS-AT linker region, and the well-documented protein-protein interactions necessary for correct function. By modelling the KS-AT dimer as a rigid body, these factors were largely ignored. The presence of additional domains within the module would be expected to impact the global structure observed. Given that enzymes often only adopted a fully ordered conformation upon substrate binding,[286] the addition of multiple catalytic domains is likely to have considerable conformational impact. The approximated oval shape of the module, and the low resolution reported, means that it is conceivable that an alternative, more convincing, model may have been generated if ensemble-state calculations had been employed, or if the structure of Dutta *et al*. had been available and used in these analyses.

A similar study, published by Davison *et al*.[144] investigating the *trans*-AT PKS virginiamycin M5, adds an additional layer of complexity. This structure, consisting of a KS dimer and two pairs of ACP domains, adopts an arch like global conformation in solution, as revealed by SAXS analysis. Although *trans*-AT PKS assembly lines are thought to have evolved independently of *cis*-AT PKSs,[287] and hence structural comparisons should be approached with scepticism, comparative analyses are still of considerable value.[288] The structure determined

by Davison *et al*. appears immediately more convincing than that of Edwards *et al*., with greater structural definition observed and supported by the docking of high resolution mPKS domain structures into the SAXS envelope (Figure 4.2d). This model reveals a global conformation which matches more closely to that reported by Dutta *et al*. It is worth noting that Edwards *et al*. were unable to generate a convincing pseudo atomic model of DEBS M3-TE by docking of domain structures into the SAXS envelope, consistent with significant dynamic heterogeneity within the module. The data obtained by Davison *et al*. reveals minimal flexibility in virginiamycin M5, despite this polypeptide structure containing two highly flexible ACP domains, attached by long linkers to the KS.

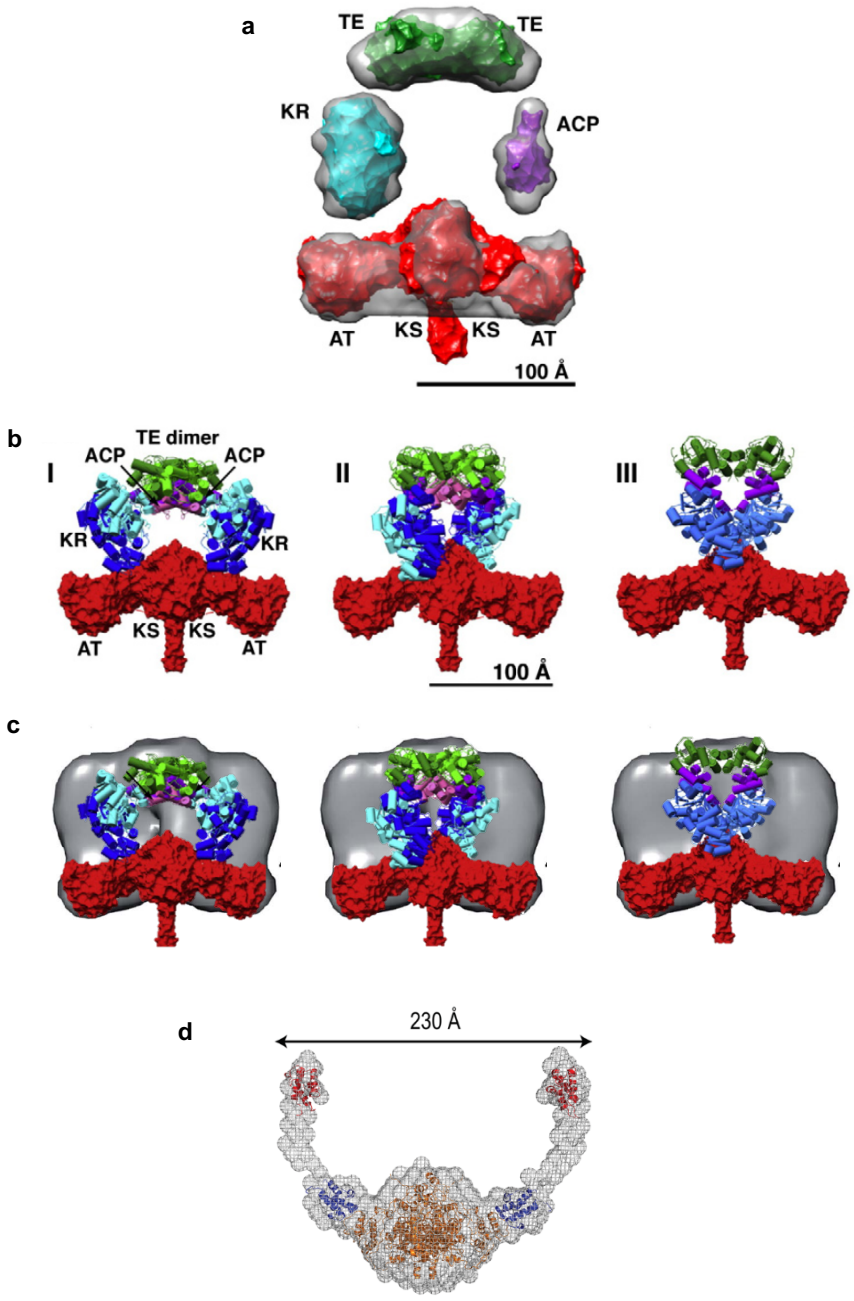



**Figure 4.2 SAXS Models of Polyketide Synthase Modules.** a) SAXS density for individual domains in DEBS M3 fits well to determined high resolution structures, although the KS didomain appears as though the slight bend towards the *pre*-KS docking domain could be caused by crystal contacts. b) The three conformations of DEBS M3+TE proposed by Edwards *et al*. c) Superposition of these conformations over the determined SAXS envelope reveals a poor fit, with a lot of unaccounted density. d) SAXS interrogation of the virginiamycin PKS M5 reveals an arch-shaped structure, with high-resolution domains fitting nicely within the density. Adapted from Edwards *et al*. and Davison *et al*.[143,144]

In addition to the studies outlined above, differences between the mammalian FAS structure determined by *Maier et al*.,[152] and mPKS structures were revealed by Zheng *et al*.[140] This study reported the first high resolution structure of an ER domain from the spinosyn PKS, which reveals a marked difference in the protein-protein interactions between the ER and KR domains, as compared to their FAS equivalents. In order to crystallise the ER domain, the team purified a didomain consisting of the KR and ER subunits. This revealed that the constituent domains adopted a side-by-side arrangement, as opposed to the end-to-end organisation seen in FAS. This significant disparity further casts doubt on the viability of using available FAS structures as a proxy for mPKSs.

### 4.1.3 Cryo-EM based PKS Module Structures

The only cryo-EM derived structure determined to date is that of PikAIII (Figure 1.7, Figure 4.3). This reveals a remarkably different domain organisation to that reported by Edwards *et al*., and is in much better agreement with that reported for M5 of the *trans*-AT virginiamycin PKS.[141] The PikAIII structure, resolved to 7.9 Å, was used as a template in which to dock high resolution crystal structures and homology models of individual mPKS yielding a complete pseudo-atomic model. Whilst not a high resolution description of an intact mPKS module, (none have been determined to date), it is a much more detailed model than that formulated by Edwards *et al*. The high resolution structure of the DEBS M5 didomain was dissected into a KS dimer and AT monomer, and complimented with atomic resolution structures of a KR domain and ACP domain, also from DEBS. The high degree of similarity necessary to create a convincing pseudo-atomic model in this way makes it unlikely that DEBS and the pikromycin PKSs adopt radically different structures in solution.

### 4.1.4 ACP Binding Sites

The ACP domain within canonical mPKS modules is universally located after the condensing and reducing domains. This means that a long flexible linker is needed for the ACP to interact with the AT, receiving its -CoA extender unit; then the KS, catalysing chain extension; and subsequently ferry the growing polyketide chain between reducing domains, before finally interacting with the downstream KS or TE. This has led to the identification of a number of possible ACP binding sites upon AT and KS domains. The structures of the KS-AT didomains

from DEBS allowed Tang *et al*. to identify an ACP docking domain situated in the cleft between the KS and AT domains.[114] Docking the NMR structure of the ACP from M3 onto the KS-AT didomain crystal structure, they showed that it was able to interact with both the AT and KS simultaneously, whilst the upstream ACP from M2[137] interacted more closely with the KS, but was shown to adopt a very similar position to the cognate ACP (Figure 4.3a). In the SAXS models constructed by Edwards *et al*.,[143] the location of the ACP could not be determined, although they positioned them diagrammatically, bound to the TE domains. In contrast, the structural model determined by Dutta *et al*.[141] shows the ACP present in two distinct locations, in a roughly 1:1 ratio. The ACP is shown to associate strongly with both the AT-KS interface and the KR domain (Figure 4.3b and c). In the arch-shaped conformation adopted by PikAIII the cleft identified by Tang *et al*. no longer exists, with the bend in the KS-AT interface obscuring this region. Consequently, the cylindrical ACP is observed to engage with the AT domain via an extended interaction interface, with the end of the α-helices presenting the key serine residue necessary to bind the methylmalonyl-CoA extender unit and the growing polyketide chain pointing inwards towards the KS dimer. In the same study, Dutta *et al*. were also able to obtain structures of PikAIII M5 complexed to the upstream ACP domain from PikAII M4. This revealed a previously unidentified ACP-KS interaction domain on the top of the arch, separate from the M5 reaction chamber. Based on the identification of these two binding sites, either side of the KS active site (Figure 4.3d and e), they proposed a hypothesis for chain transfer between modules. This configuration provides a clear structural basis for the principal of vectoral biosynthesis mPKS assembly lines, where ACPs from different modules would never bind in the same location, allowing the KS to clearly discriminate between the two. Intriguingly, the ACP from M4 would only bind to the KS of M5 when a polyketide analogue, similar to the natural growing chain, was bound. These data imply that the presence of substrate drives conformational changes necessary for protein-protein interaction. Thus, to structurally observe any *inter*-module interactions in mPKS systems, polyketide analogues will need to be synthesised and loaded onto relevant ACP domains.

In a companion paper, Whicher *et al*. were able to determine mid-resolution cryo-EM structures for PikAIII M5 locked into a series of different states within the catalytic cycle.[142] These were accessed by incubating PikAIII M5 with different substrates and analogues thereof. The resulting structures identified a number of crucial structural rearrangements necessary for catalysis to proceed. With the natural pentaketide substrate bound to the KS dimer *via* an active site cysteine residue, the KS undergoes a distinctive conformational change. This is

hypothesised to lead to a slight shift in the AT, obscuring the side entrance through which the upstream ACP delivers the polyketide chain. The ACP is brought into close contact with the AT domain, and appears to be conformationally aligned to accept an AT-loaded methylmalonyl extender unit (Figure 4.3f). The largest conformational change is reserved for the KR domain, which flips end-to-end in a roughly 180° rotation. This large conformational change, precipitated by substrate binding, is another example for why rigid body fitting should be avoided wherever possible when modelling mPKS structures. Chain extension leads to the ACP docking close to the cleft between the AT and KR domains (Figure 4.3g). With the natural reduced hexaketide substrate bound to the ACP, a further conformational change is observed. The ACP now re-locates to one of three positions beneath the KR and is ejected out of the module reaction chamber (Figure 4.3h). This would allow it to interact with the downstream module, forming a transient interaction with the apex of the M6 KS domain, passing the elongated polyketide chain onto this neighbouring module. Presumably the module would then return to one of the conformations observed by Dutta *et al*., revealing the active site channel on top of the arch, now primed to receive a new pentaketide substrate.

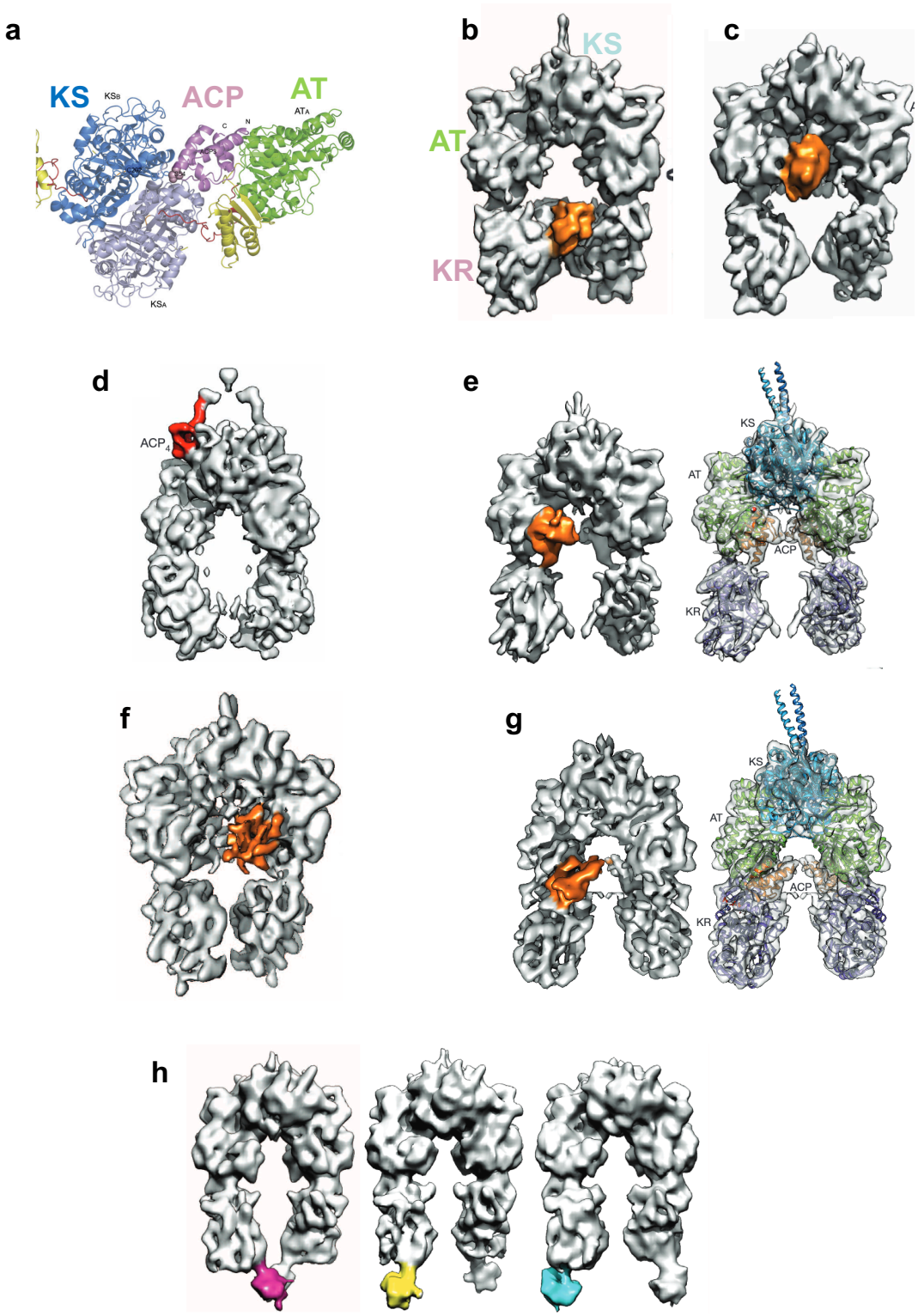



**Figure 4.3 ACP Location and Movement in Different PKS Models.** a) ACP docking site determined by Tang *et al*. The NMR structure of DEBS M2 ACP[137] was used to create a homology model for DEBS M3 ACP (purple). This was docked into the X-ray crystal structure of DEBS M3 AT-KS didomain. b) and c) Cryo-EM reconstructions of PikAIII showing the two ACP locations (orange)

when no substrates are bound. d) Cryo-EM reconstruction highlighting the docking site for the upstream ACP from PikAII M4 (red), on top of the arch shaped structure. This structure does not contain the ACP from pikromycin M5. e) Cryo-EM reconstruction of PikAIII with its cognate ACP bound to the AT, ready to receive a methylmalonyl extender unit. f) Cryo-EM reconstruction showing the location of methylmalonyl-loaded PikAIII M5 ACP at the bottom entrance to the KS active site, ready to receive the pentaketide substrate for chain elongation. g) Cryo-EM reconstruction of PikAIII with an unreduced hexaketide substrate loaded onto the ACP, which is bound to the KR for NADH coupled reduction. h) Cryo-EM reconstructions of the three conformations observed when the ACP of PikAIII is loaded with a fully extended and reduced hexaketide product, ready for transfer to PikAIV M6. Figure adapted from Tang *et al*., Dutta *et al*. and Whicher *et al*.[114,141,142]

### 4.1.5 The Turnstile Mechanism of Vectoral Synthesis

The cryo-EM structures of PikAIII provide a compelling structural basis for the high-fidelity vectoral synthesis employed by mPKSs. Lowry *et al*.[146] provide convincing biochemical evidence for the ACP movement observed by Whicher *et al*., though due to the two competing models for polyketide synthase architecture, there is no mention of either study by Skiniotis and co-workers.[141,142] The differences observed between the structures of FASs and the PikAIII cryo-EM models could be a result of the need to ensure high fidelity vectoral synthesis in multimodular PKSs. The explicit need to pass a growing polyketide chain between modules, without skipping or premature termination, may be achieved by sequestering each round of chain elongation and reduction within its own catalytic chamber, as formed by the arch shaped conformation. In FAS systems, which do not require the same level of fidelity as they are not multimodular in nature,[289] a more open configuration of the complex is tolerated due the reduced requirement to precisely control protein-protein interactions. One potential benefit of the relaxation of high-fidelity regulation is an increase in rate of biosynthesis.

The structural information also provides rationale for Keatinge-Clay's proposal to redefine the modular architecture.[169] Without the highly specific protein-protein interaction between the preceding ACP and the downstream KS, the PKS assembly line would stall, locking modules in a 'closed' conformation, with the active site tunnel on top of each KS occluded by the AT domain. Within the enclosed reaction chamber, the specificity required for the protein-protein interaction between the ACP and the KS, AT or KR is likely to be decreased, as ACPs from other modules will be occluded from binding by their exclusion from the reaction chamber formed by the arch-shaped configuration of the AT and KR domains.

### 4.1.6 The Aim of this Work

The function of mPKSs is predicated upon precise protein-protein interactions between modules, both *inter*- and *intra*-polypeptide.[91] Attempts to engineer these systems have proved challenging due to a lack of detailed understanding about the nature of the interactions necessary for correct function.[130,131] The model derived by cryo-EM does not account for the crystal structures of the KS-AT didomains,[113,114] which have been used to derive a FAS-like architecture of mPKS modules. This itself does not account for the convincing cryo-EM reconstructions of PikAIII.[141,142] The competing structural models proposed for bacterial mPKS architectures are likely to confuse attempts to rationally engineer these assembly lines as different protein-protein interactions are identified.[170] They also lead to different mechanistic interpretations and different models for *inter*-module communication. The work outlined in this chapter seeks to resolve the apparent disparities between these conflicting models using a hybrid structural biology approach combining X-ray crystallography and cryo-EM. It is hoped that by adopting this approach a more comprehensive model for mPKS architecture can be obtained. The test subject chosen for these studies was the abyssomicin C synthase minimal module AbyB3.

## 4.2 Methods

### 4.2.1 Expression and Purification of AbyB3 and AbyB3ΔACP

AbyB3 was cloned and expressed according to the method outlined in Chapter 3. A small modification was made to the purification protocol, which resulted in enhanced enzyme stability. During purification by SEC, AbyB3 was buffer exchanged into XL Buffer (100 mM NaP, 150 mM NaCl, 1 mM EDTA, 0.02% sodium azide, pH 7.5). If a sample was required for NSEM, this material was further buffer exchanged into 50 mM HEPES pH 8.0 prior to use. AbyB3ΔACP was cloned as described in Section 4.3.1. This polypeptide was expressed and purified as for AbyB3 with minor modifications as follows: i) HEPES buffers were used throughout all purification steps (see Section 3.2.4). ii) Each buffer (Load, Elute and SEC) was additionally supplemented with 10 % glycerol to improve the stability of the protein. AbyB3ΔACP was either used immediately or flash frozen in liquid nitrogen and stored at -70 °C prior to use.

### 4.2.2 Crystallisation

Recombinant proteins were purified as described above, following which they were concentrated to 10 mg/mL using a vivaspin centrifugal concentrator (Sartorius) with a 50 kDa MW cut-off membrane. Concentrated proteins were centrifuged at 19,000 x g for 5 minutes to remove any aggregates prior to crystallisation screening. Crystallisation experiments were conducted using commercially sourced screens (PACT premier, Structure Screen 1&2, JCSG-plus, MIDASplus, Morpheus and LMB crystallization screen; all from Molecular Dimensions). Screens were established using 50 µL of each crystallisation condition placed into the reservoir of a 96 well crystallisation plate. Two drops per reservoir were established to enable crystallisation via the sitting drop vapour diffusion method. One drop comprised 0.5 µL reservoir condition and 0.5 µL 10 mg/mL protein, whilst the second comprised 0.75 µL reservoir condition and 0.25 µL protein. Plates were stored at 20 °C and monitored periodically for a period of 12 months. Diffraction quality crystals of AbyB3ΔACP were grown over a period of two weeks in 0.1 M MES, 10 % ethanol, 30 % PAA 5100, pH 6, using 0.5 µL of reservoir and 0.5 µL of 10 mg/mL AbyB3ΔACP.

### 4.2.3 X-Ray Diffraction Data Collection and Processing

Crystals grown in the condition described above were manually looped following cryo-protection by the addition of 25 % glycerol to the sitting drop. They were flash-frozen in liquid nitrogen and subjected to X-ray diffraction analysis at Diamond Light Source (Beamline I03). Four crystals were interrogated, with the best diffracting to a maximum resolution of 2.36 Å.

Diffraction data were initially processed on-site, using xia2 dials, and determined to be in the space group C1 2 1. The structure was subsequently determined using by molecular replacement using the program BALBES,[290] employing elements of both DEBS didomains and mammalian FAS structures as search models.[113,114,136,152] Electron density maps were examined and determined to be of good quality. The structure was further refined using the CCP4 suite of programs,[291] employing multiple iterative rounds of model building in Coot[292] and refinement in REFMAC5. The residues 0-32, corresponding to the *pre*-KS docking domain and residues 888 – 915, corresponding to part of the *post*-AT linker, were not resolvable due to insufficient electron density. Residues 280, 457 – 465 and 685 – 694, corresponding to flexible loops within the structure, remain unresolved for the same reason. A full complement of water molecules and ions resolvable within the structure has not yet been determined.

### 4.2.3 Crystal Structure Analysis

The structure of AbyB3ΔACP was analysed using PyMOL.[293] Inbuilt functions were used to determine the distance between atoms for hydrogen bond and salt bridge identification, and for $C_\alpha$ rmsd calculations.

### 4.2.4 Crosslinking AbyB3

A variety of commercial crosslinking methods were applied to recombinant AbyB3. Each method was performed according to the supplier's instructions. The most effective crosslinking method, using $PEG_{12}$-SPDP (ThermoFisher Scientific), was carried using the following protocol. AbyB3 was purified in XL buffer as previously described. It was then immediately subjected to chemical crosslinking. A 20 mM solution of $PEG_{12}$-SPDP was prepared and

immediately added to AbyB3 at 5 mg/mL, to give a final concentration of 500 μM. Reactions were subsequently incubated at room temperature for 60 minutes. The resulting cross-linked material was then passed through a Zeba Spin Desalting column (ThermoFisher Scientific), pre-equilibrated with XL buffer. Half of the conjugated AbyB3 was placed on ice, and the other half was reduced by the addition of DTT dissolved in acetate buffer (100 mM sodium acetate, 100 mM NaCl, pH 4.5) to a final concentration of 50 mM. This sample was incubated for 30 minutes before application to a Zeba Spin Desalting column, pre-equilibrated in XL buffer, to remove DTT. The two protein solutions were subsequently mixed and left at 4 °C for 18 h.

### 4.2.5 Preparation of Cryo-EM Grids

Cryo-EM grids were prepared within 30 hours of *E. coli* cell lysis to ensure the sample was as fresh as possible. Further details are given in Section 4.4.2. For clarity this method is described as follows. Immediately prior to use, grids (apart from those coated in graphene oxide) were glow-discharged in air at a pressure of 6 x $10^{-3}$ atmospheres using a current of ~15 mA for 1 minute. Either a Leica ACE 600 (Leica Microsystems) or an ELMO Glow Discharge System (Agar Scientific) was used. A plunge freezer, either a Leica GP (Leica Microsystems) or Vitrobot Mk IV (FEI), was pre-cooled to < -170 °C using liquid nitrogen. Liquid ethane was subsequently added to fill the cryogen holder. The sample application chamber was typically cooled to 4 °C and an atmosphere of 100 % humidity formed. 4 μL of AbyB3 in XL or HEPES buffer, typically between 0.1 and 0.2 mg/mL, was applied directly to the grid, which was suspended within the sample application chamber. The grid was then blotted and immediately plunged into liquid ethane to ensure vitrification. Grids were subsequently transferred into a grid box (Electron Microscopy Sciences) maintained under liquid nitrogen. Grids were then either stored in liquid nitrogen until use, or immediately taken forwards for clipping. The grids have an inner and outer ring applied by hand, so that they can be manipulated by the autoloading system within the Talos Arctica microscope (FEI). Once clipped, they are either stored or transferred directly into a sample puck for insertion into the microscope and subsequent analysis.

### 4.2.6 Cryo-EM Grid Screening

Cryo-EM grids were typically screened using a Talos Arctica 200 kV microscope (FEI) equipped with a Gatan K2 direct electron detector. The microscope was prepared using both direct alignments and software optimisations. Initially, an Atlas image of the grid is acquired, before a suitable gridsquare is selected. A eucentric height is then acquired and image shifts determined. Quantifoil holes to be imaged were subsequently selected, or a pattern created for Lacey Carbon grids. Before data acquisition, direct alignments of the microscope were performed to ensure image quality. The C2 aperture was aligned and stigmated before the beam tilts and shifts were adjusted. The diffraction pattern was subsequently focussed, and images automatically acquired by the EPU software (FEI). In order to screen each grid, a few micrographs were obtained from a sub-set of gridsquares exhibiting variations in ice thickness. The micrographs obtained were then assessed for particle quality and dispersion using Relion 3.0.[294]

### 4.2.7 Cryo-EM Data Collection

Data collection was performed using Lacey Carbon grids found to contain a good concentration and dispersion of particles during screening. The microscope was prepared using a similar protocol, but ensuring the conditions outlined in Table 4.4 were specified. 7 gridsquares were imaged overnight, collecting a total of 160 micrographs. 138 of these were judged to have an adequate contrast transfer function (CTF) as calculated using Relion 3.0, and were taken forward for manual processing using Relion 3.0.

## 4.3 Results: Structural Interrogation of AbyB3 by X-Ray Crystallography

To elucidate the structure of the intact AbyB3 module and address current disparities in published mPKS structural modules, two primary research objectives were formulated. The first of these was to obtain high-resolution structural data of each AbyB3 domain and the second was to elucidate the native structure of intact AbyB3.

The best technique to address the first of these priorities is X-ray crystallography. Due to the inherent challenges associated with crystallising intact PKS modules, obtaining crystals of a full length AbyB3 module was considered to be unlikely. With this in mind, an AbyB3ΔACP construct was generated, consisting of the *pre*-KS docking domain, the KS, the KS-AT linker and the AT-ACP linker (Figure 4.4a). This construct should not only have a higher chance of crystallisation, due to the absence of the highly dynamic ACP domains, but would also provide the opportunity to make a direct comparison with the DEBS didomain structures reported previously.[113,114] This would allow us to determine whether the mammalian FAS homology model developed by Khosla and co-workers could also be applicable to bacterial mPKSs beyond DEBS.

### 4.3.1 Cloning and Expression of AbyB3ΔACP

AbyB3ΔACP was initially cloned from a pOPINF plasmid containing AbyB3 genomic DNA, using *AbyB3_pE_Fwd* and *AbyB3_pE_Rvs* primers (Appendix Table A1). PCR amplification was completed in one step (Figure 4.4). The construct was inserted into the plasmid pOPINE, which contains an ampicillin resistance marker, akin to pOPINF, but differs in that it possesses a much shorter, C-terminal, hexa-histidine tag for purification by nickel affinity chromatography (Appendix Table A2). This approach was taken in an effort to minimise the impact of flexible linkers and maximise the likelihood of successful crystallisation.

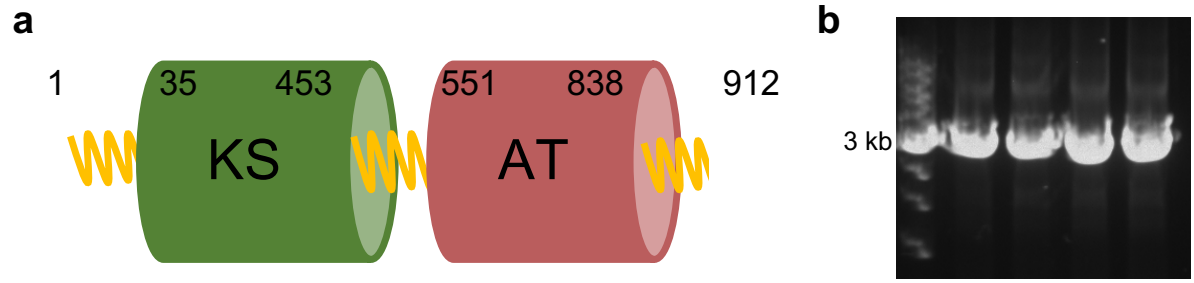



**Figure 4.4 Cloning AbyB3ΔACP.** a) Schematic Representation of the Abyssomicin PKS M7 AT-KS Didomain. A truncated version of AbyB3 was created, including the *pre*-KS docking domain, the KS, the KS-AT linker, the AT and the AT-ACP linker. b) PCR amplification of AbyB3ΔACP. The ~2.7 kb gene was amplified using a gradient PCR with temperatures ranging from 60°C to 65°C during the annealing step.

AbyB3ΔACP::pOPINE was transformed into both *E. coli* BL21 (DE3) and *E. coli* BL21 BAP1 cells for expression trials. Although expression was observed in both cell lines, a notable increase in expression levels were observed in BAP1 cells. For this reason, BAP1 cells were chosen as the cell line of choice for large scale purification attempts. Although AbyB3ΔACP expression was observed at high yields, this polypeptide was found to be much less soluble in buffer than AbyB3, with high levels of protein aggregation observed following purification by SEC. The addition of 10 % glycerol to the purification buffers used in the production of this protein ensured continued solubility and enabled the production of AbyB3ΔACP to >95 % purity (Figure 4.5).

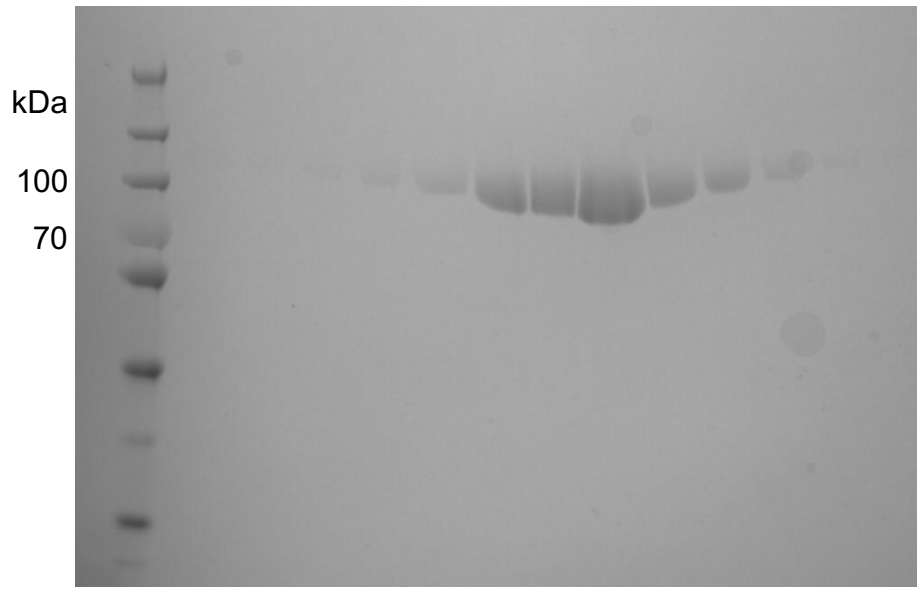



**Figure 4.5 Expression of AbyB3ΔACP.** An SDS-PAGE gel of purified AbyB3ΔACP. The construct was purified using nickel affinity chromatography and subsequent SEC, all in the presence of 10 % glycerol to ensure solubility.

### 4.3.2 Crystallisation and X-ray Diffraction Studies of AbyB3ΔACP

Purified AbyB3 and AbyB3ΔACP, at a concentration of 10 mg/mL, was subjected to crystallisation screening using six commercially available 96-condition screens, employing the sitting drop method of vapour diffusion, as described in Section 4.2.2. Crystallisation attempts for AbyB3 yielded no viable crystals. However, a single condition from the MIDASplus screen returned small crystals of AbyB3ΔACP (Figure 4.6a), which formed over a period of four weeks. This condition comprised 0.1 M MES, 10% ethanol, 30% PAA 5100, pH 6. Several of the crystals present in the drop were subjected to X-ray diffraction at Diamond Light Source. Data collection and subsequent processing was performed as described in Section 4.2.3. The data collection statistics and final refinement statistics determined to date are displayed in table 4.1.

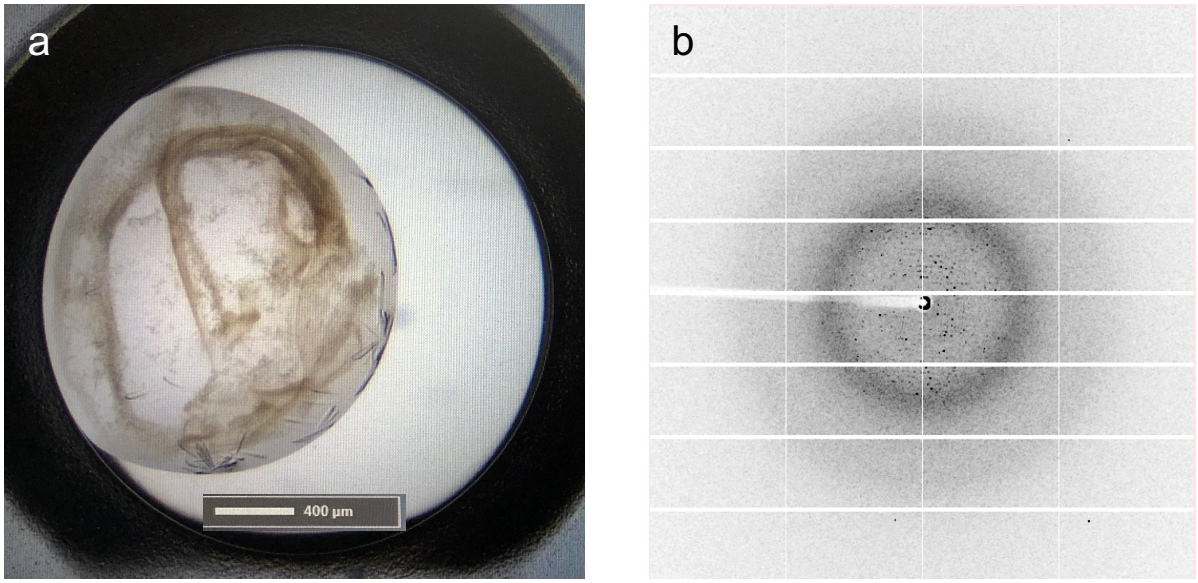



**Figure 4.6 Crystallisation and X-Ray Diffraction of AbyB3ΔACP.** a) Rectangular crystals of AbyB3ΔACP grown in 0.1 M MES, 10% ethanol, 30% PAA 5100, pH 6. b) X-ray diffraction of AbyB3ΔACP to a maximum resolution of 2.36 Å at Diamond Light Source, beamline i03.

| Data Collection Statistics | |
|---|---|
| Wavelength (Å) | 0.98 |
| Resolution (Å) | 2.36 |
| Reflection (observed / unique) | 200817 / 28752 |
| Completeness (%) | 100 (100) |
| Multiplicity | 7.0 (6.5) |
| I/σ(I) | 3.4 (0.9) |
| $R_{merge}$ | 0.544 (2.474) |

| Refinement Statistics | |
|---|---|
| Space Group | *C*121 |
| Unit Cell Parameters (Å) | a = 165.3, b = 86.2, c = 65.7 |
| Monomers per A.U. | 2 |
| Protein Atoms | 11,823 |
| *R*-factor | 0.246 |
| *R*-free | 0.321 |
| RMS Deviations | |
| Bonds (Å) | 0.0070 |
| Angles (°) | 1.69 |
| Ramachandran Plot (%) | |
| Favoured | 95.2 |
| Allowed | 3.9 |
| Disallowed | 1.0 |

**Table 4.1 Diffraction and Refinement Data for AbyB3ΔACP.** Statistics in parenthesis are for the highest resolution shell. Rmerge= $\sum |I_o - \langle I \rangle| / \sum \langle I \rangle$ summed over all observations and reflections. Additional cycles of model building and refinement were conducted by Dr. Nicholas Burton at the University of Bristol.

### 4.3.3 X-ray Crystal Structure of AbyB3ΔACP

Although the final stages of model building and refinement of the AbyB3ΔACP are ongoing towards a completed crystal structure of this polypeptide; presented herein is our current lead model of AbyB3ΔACP. The remaining required minor modifications of this structure include the introduction of alternate conformers and revision of a small number of regions which are poorly defined in the electron density map. The structure of AbyB3ΔACP has been determined

to 2.36 Å resolution, with our current model well suited for analysis and interpretation (Figure 4.7). The KS and AT domains of AbyB3ΔACP are both well defined, as is the KS-AT linker region. The KS domains dimerise across a large interface with the KS-AT linker and the AT domain stretching horizontally outwards in opposite directions. The AT-ACP linker is partially defined, with the final 24 amino acids showing no corresponding density. This region appears to wrap around the KS-AT linker and attaches to the bottom of the KS domain, close to the ACP docking domain, as identified by Whicher *et al.* in their cryo-EM structure.[142] Unfortunately, no density corresponding to the *pre*-KS docking domain is visible, indicating flexibility in this region.

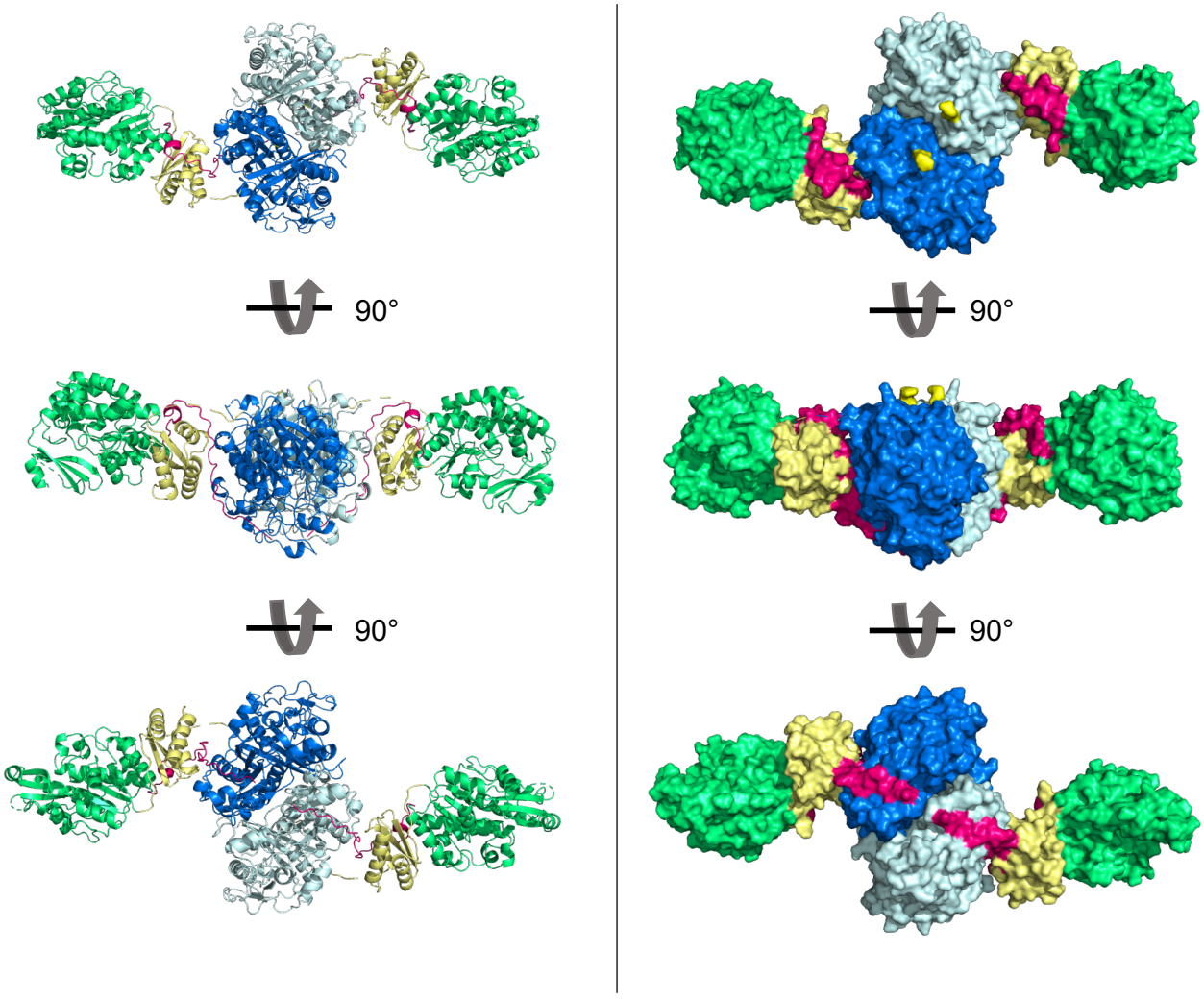



**Figure 4.7 The X-Ray Crystal Structure of AbyB3ΔACP.** The structure of dimeric AbyB3ΔACP is depicted using a cartoon representation (left) and a surface representation (right). One KS subunit is shown in blue with the other in cyan. Both AT domains are shown in green, with the KS-

AT linker regions in light yellow. The one residue distinguishable from the *pre*-KS docking domain is shown in bright yellow and the AT-ACP linker is shown in pink.

### 4.3.3 Comparison with the Structures of DEBS M3 and M5

The crystal structure of AbyB3ΔACP is immediately comparable to those of the DEBS M3 and M5 didomain structures. Each adopts a similar global fold, which can be readily superimposed (Figure 4.8). This analysis reveals small $C_\alpha$ rmsds between the atoms present in each structure. AbyB3ΔACP is more closely structurally related to DEBS M5 with a 3.4 Å $C_\alpha$ rmsd, while a larger value of 7.3 Å $C_\alpha$ rmsd is determined between AbyB3ΔACP and DEBS M3. Interestingly, AbyB3ΔACP can thus be considered to possess greater structural identity to DEBS M5 than DEBS M3 (5.1 Å $C_\alpha$ rmsd). Whilst the sequence identities between AbyB3ΔACP and DEBS M3 and M5 are high, 45 % and 48 % respectively, these values are lower than between DEBS M3 and M5 (54 %). This serves to highlight the structural conservation between mPKS modules. The major identified differences between these structures are attributable to the 27° change in angle between DEBS M3, and both DEBS M5 and AbyB3ΔACP, as previously postulated by Tang *et al*.[114] This is caused by a combination of crystal packing constraints and the removal of the *pre*-KS docking domain, which was observed to mildly inhibit function in DEBS M3.

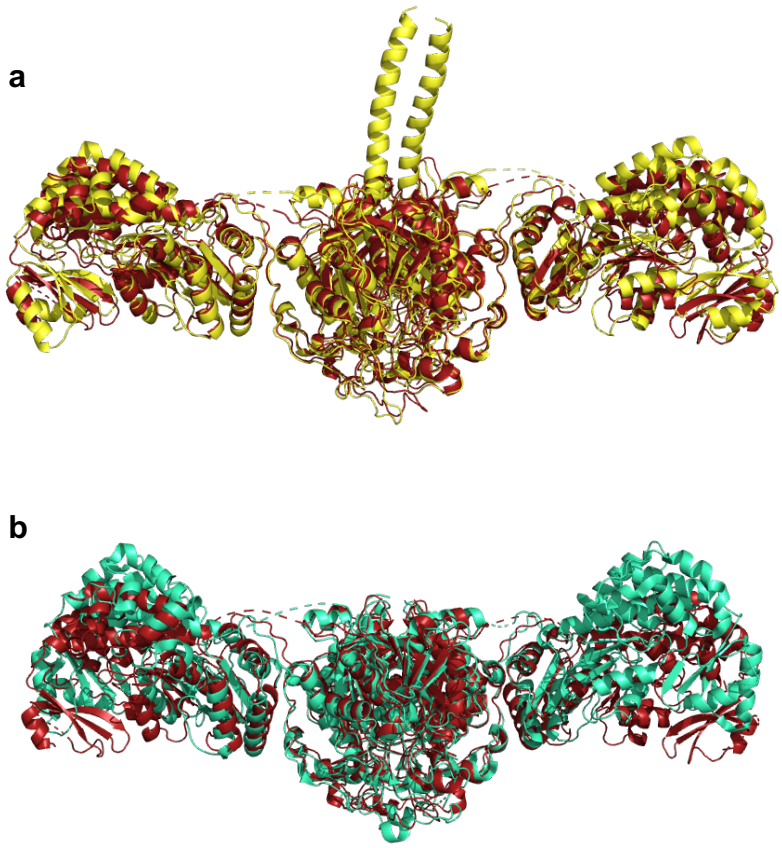



**Figure 4.8 Comparison of AbyB3ΔACP with Didomain Structures from DEBS.** The structure of AbyB3ΔACP (red) overlaid with the structure of a) DEBS M5 (yellow) and b) DEBS M3 (teal).

### 4.3.4 Crystal Constraints Imposed on AbyB3ΔACP

In order to investigate these proposals further, the crystal packing arrangement observed for AbyB3ΔACP was explored. As with the structure of the DEBS M5 AT-KS didomain, significant *inter*-dimer contacts are observed (Figure 4.9a). With each dimer forming an 'S' conformation, individual protomers are able to stack closely together, potentially stabilising the linear structure. These sheets of interlocking 'S' conformations do not, however, stack directly on top of each other. Instead, the AT domains from dimers in the plane above sit in the cleft created by the KS-AT linker (Figure 4.9b). This results in the formation of a cavity above the KS domains, in between the AT domains from the dimers on the above plane. This provides free space for the flexible *pre*-KS docking domains to be accommodated, though the absence of compelling electron density for these regions points to their highly flexible nature. These observations

are in contrast to the packing constraints imposed in the crystallisation of DEBS M5, where the *pre*-KS docking domain sits in the cleft created by the KS-AT linker region, thereby stabilising it and allowing its resolution (Figure 4.1). Whilst the presence of significant packing constraints does not preclude the *pre*-KS linker having a significant effect on the structure of a didomain, it does show that both hypotheses have merit.

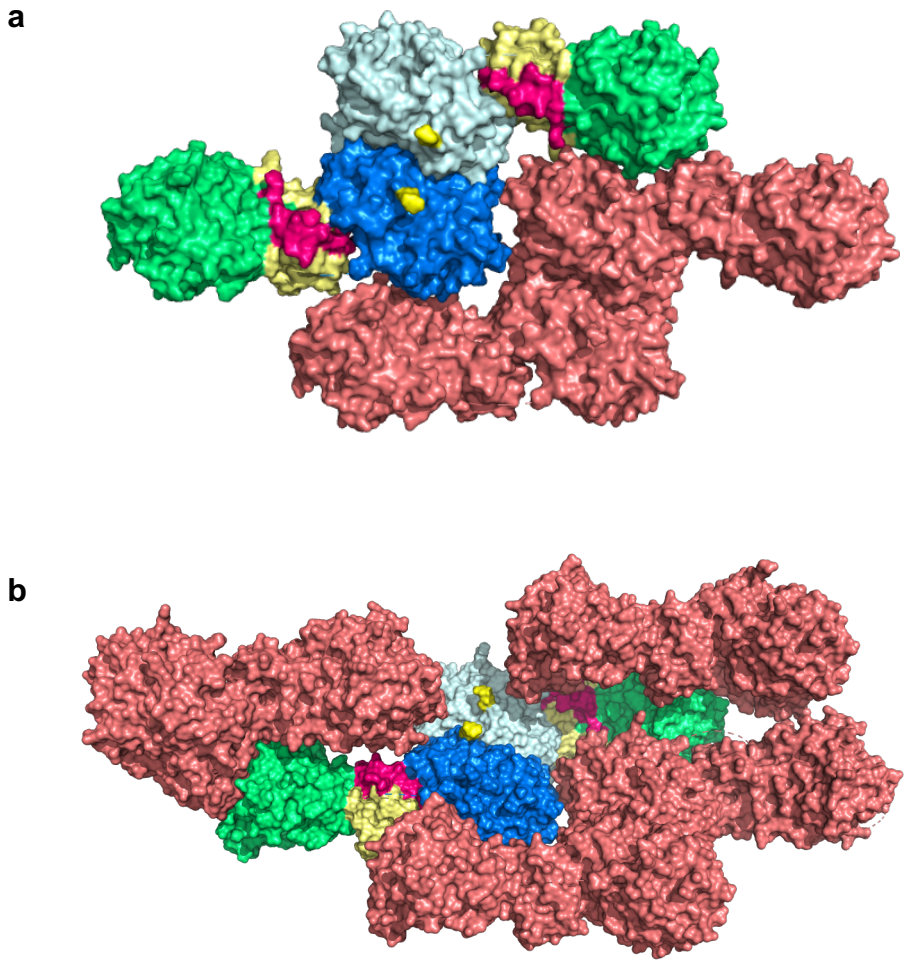



**Figure 4.9 Crystal Packing Constraints for AbyB3ΔACP.** Each AbyB3ΔACP dimer forms significant contacts with other dimers within the crystalline lattice. a) On a single plane the dimers stack together, with the AT domain from one dimer resting between the KS and AT of another dimer. b) These contacts also exist both above and below the dimer. Shown are two monomers which illustrate the cavity in which the *pre*-KS docking domain sits unconstrained.

### 4.3.5 Analysis of Individual Domains in AbyB3ΔACP

The KS domain of AbyB3ΔACP adopts a five-layered αβαβα fold, as seen in previous KS structures, including those from type I and type II PKSs, and *trans*-AT PKSs, as well as in homologous condensing domains from FASs.[99,113,114,295,296] The KS dimerisation interface is extensive, stretching 50.4 Å from corner to corner and is dominated by hydrophobic interactions and hydrogen bonds. The core of each KS monomer is composed of five antiparallel β-sheets. Across the dimerisation domain the two five-membered β-sheets are joined by backbone hydrogen bonds, creating a ten-membered anti-parallel β-sheet structure (Figure 4.10a). Whilst this is a common feature to previously studied KS domains,[99,295,297] specificity is conferred by two salt bridges between arginine 190 and glutamic acid 313 (Figure 4.10b). This pair of charged residues reside at the top two corners of the dimer interface, a notable difference from DEBS M3, which lacks salt bridges, and DEBS M5, where two salt bridges form at the base of the dimerisation interface. The conservation between the structure of distinct KS domains is demonstrated by a 1.31 Å rmsd between all $C_\alpha$ atoms in the DEBS M5 and AbyB3 KS domains, decreasing to 0.56 Å once outliers have been removed.

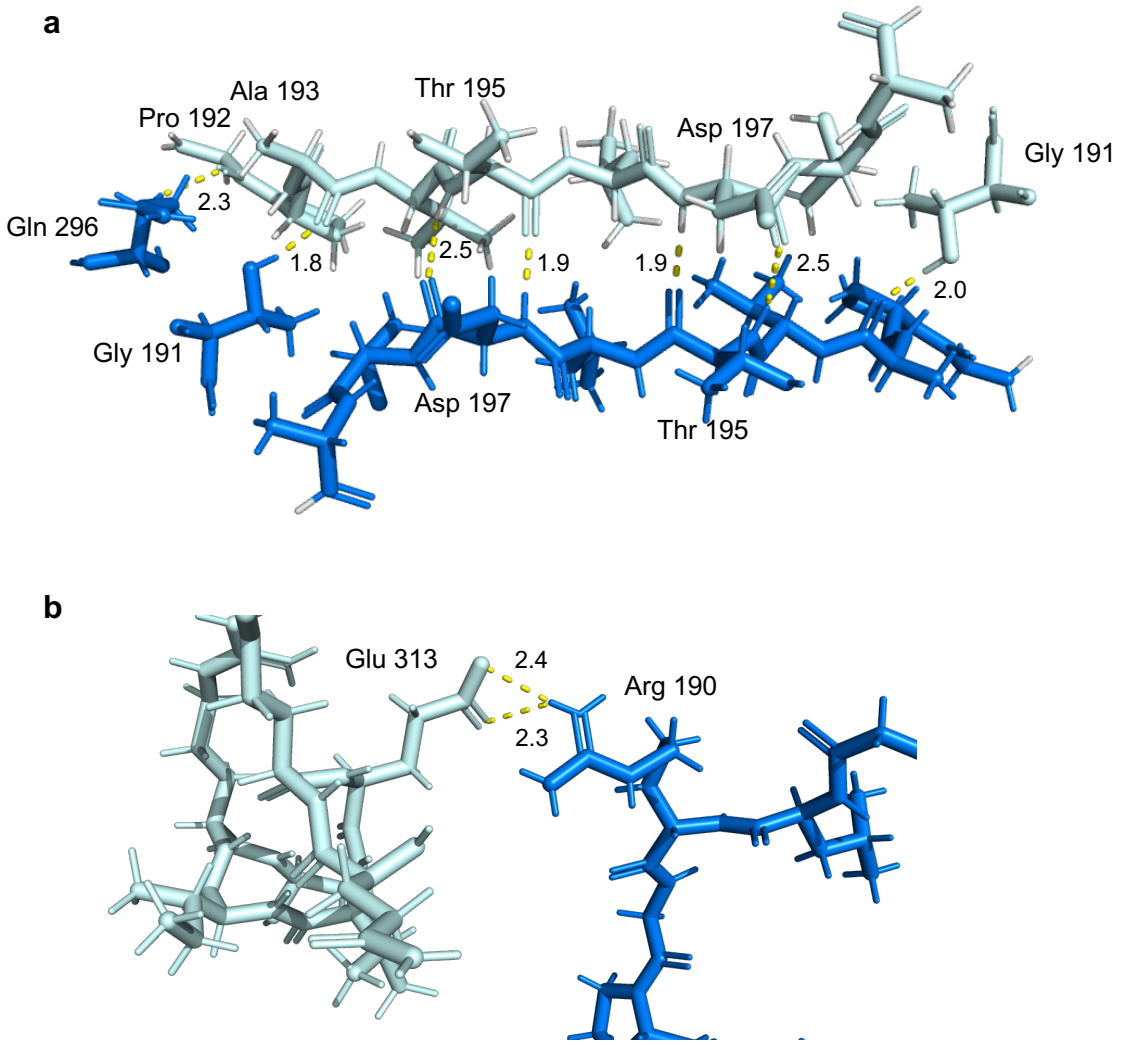



**Figure 4.10 Identification of the KS Dimerisation Domains.** a) Identification of backbone hydrogen bonds formed between two antiparallel β-sheets in opposing monomers, as well as a couple of surrounding hydrogen bonds. b) The salt bridge formed between R190 and D313. These salt bridges are formed at the top two corners of the dimerisation interface. Distances given are in Angstroms.

The active site in both the KS and AT domains of AbyB3ΔACP can be identified using both sequence and structural homology to published didomain structures. The active site in the KS domain houses a highly conserved catalytic triad consisting of cysteine 200, histidine 335 and histidine 373 (Figure 4.11a). This triad faces downwards and towards the dimerisation interface. Whilst Cys200 binds the polyketide chain, the two histidine residues have been shown to play a key role in the decarboxylation of the ACP-bound extender unit.[298] The AT domain consists of a ferredoxin-like fold fused to a larger α/β-hydrolase sub-domain, that forms the central core of the AT (Figure 4.11b). The active site resides at the interface of these two structural

elements, and houses the proposed critical residue serine 637. Whilst the structural identity between AbyB3 AT and DEBS M5 AT is high, with a 3.4 Å rmsd between $C_\alpha$ atoms, these domains are noticeably more structurally differentiated than their respective KS domain partners.

The KS-AT linker region of AbyB3ΔACP closely mirrors that of both DEBS didomains, with two anti-parallel β-sheets sitting between three α-helices (Figure 4.11c). The β-sheets are flanked on one side by two additional α-helices in the DEBS structures, one belonging to the AT and one to the AT-ACP linker. In the AbyB3ΔACP structure the α-helix belonging to the AT-ACP linker is only comprised of one full turn, raising the possibility that, when the ACP domain is attached, this helix could dissociate into a flexible loop, enabling a conformational rearrangement to bring the AT domains down into the arch conformation observed by *Dutta et al*.[141]

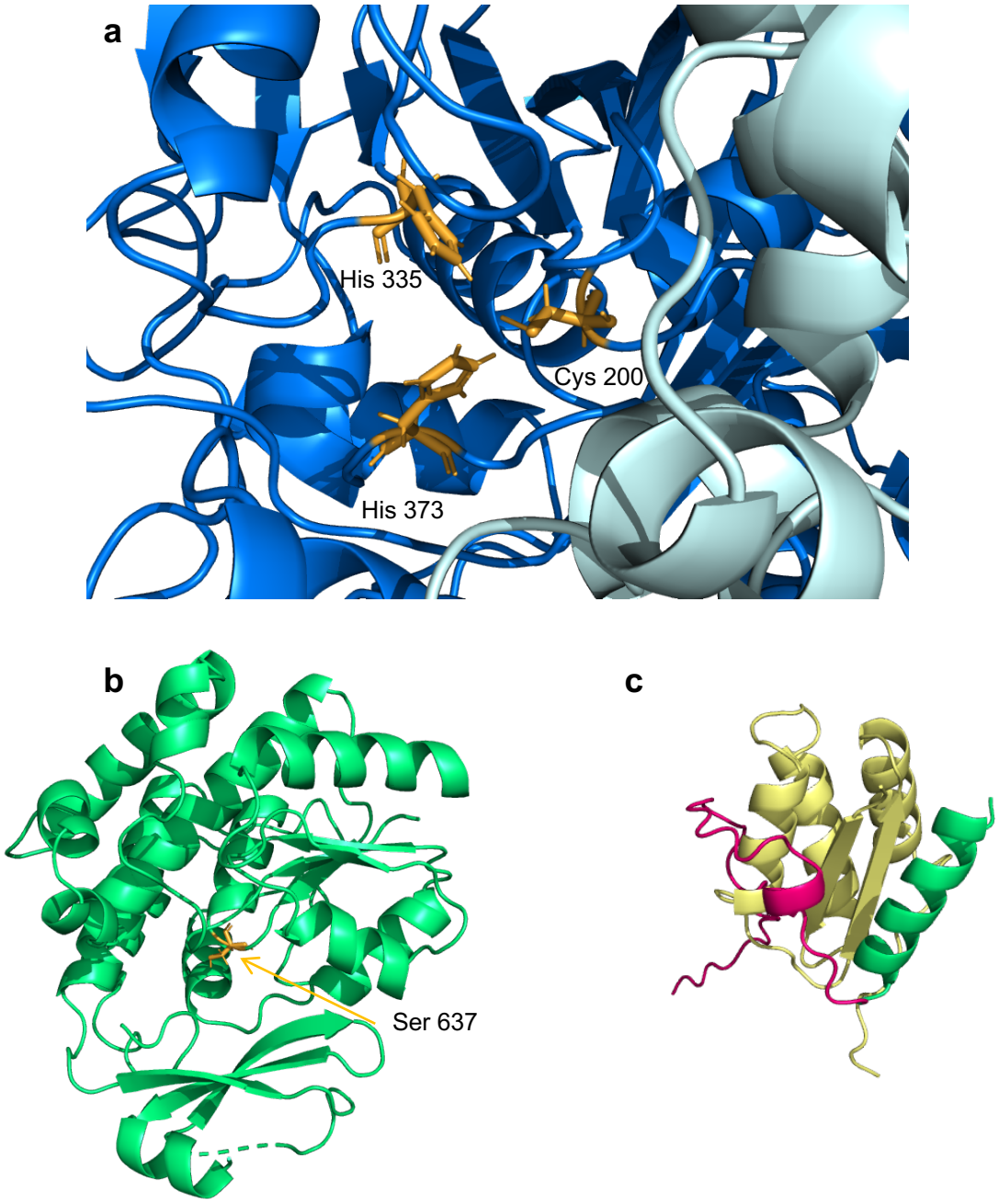



**Figure 4.11 Visualising the KS and AT Domain Active Sites and the KS-AT Linker Region in Detail.** a) The KS active site is highlighted, with the catalytic triad of Cys200, His335 and His373 shown in orange. b) The AT domain with the catalytic Ser637 residue highlighted in orange. The ferredoxin-like fold resides at the bottom of the image. c) The KS-AT linker region is shown in light yellow, with an interacting α-helix belonging to the AT domain shown in green, and the loop of the AT-ACP linker shown in pink.

## 4.4 Results: Structural Analysis of AbyB3 by EM and Cryo-EM

In order to determine the structure of the intact minimal module of AbyB3, thereby investigating what conformational constraints are imposed on the linear didomain structure by the addition of the ACP domain, single particle cryo-EM was used. The multiple structures obtained by Dutta *et al*. and Whicher *et al*.[141,142] leave significant room for resolution improvement and provide limited information about what an intact minimal module might look like. The previously determined DEBS didomain structures[113,114] are far closer in terms of sequence to a minimal module than PikAIII. The 90 amino acids that comprise the AbyB3 ACP domain represent roughly 10 % of the intact polypeptide, whereas an additional KR domain, and linker, as is present in PikAIII, equates to a > 30 % increase in module molecular weight.

### 4.4.1 Crosslinking AbyB3

In an attempt to obtain as high a resolution structure of AbyB3 as possible, a number of crosslinking methods were investigated to bias the formation of the AbyB3 tetramer seen in solution (Figure 3.9). This work was undertaken as, when solving cryo-EM structures, larger proteins are easier to work with and deliver higher contrast in the resulting image, leading to increased resolution. Additional axes of symmetry also enable structures to be solved at a higher resolution. This work was to be undertaken whilst simultaneously attempting to solve a structure of the modular dimer, as shown in the studies of PikAIII and DEBS modules.[91,114,141–143]

Although multiple crosslinkers were investigated, including BS3, glutaraldehyde (both linking two primary amine groups) and Sulfo-SBED (linking a primary amine with any other amino acid functional group), the best results were seen when crosslinking AbyB3 with $PEG_{12}$-SPDP, a multifunctional crosslinker with a 54.1 Å linker length (Figure 4.12). This required binding $PEG_{12}$-SPDP to AbyB3 via a primary amine, before reducing the disulphide bond in the linker with DTT. This reveals an activated sulfhydryl group that could either bind directly to other sulfhydryl groups or link to another activated subunit of AbyB3, hence fusing two primary amines *via* a 108.2 Å linker. Although $PEG_{12}$-SPDP is a long, non-specific linker, this was deemed preferable to generating new constructs capable of being linked specifically, both to save time and because nothing is known about how the formation of a tetramer might occur.

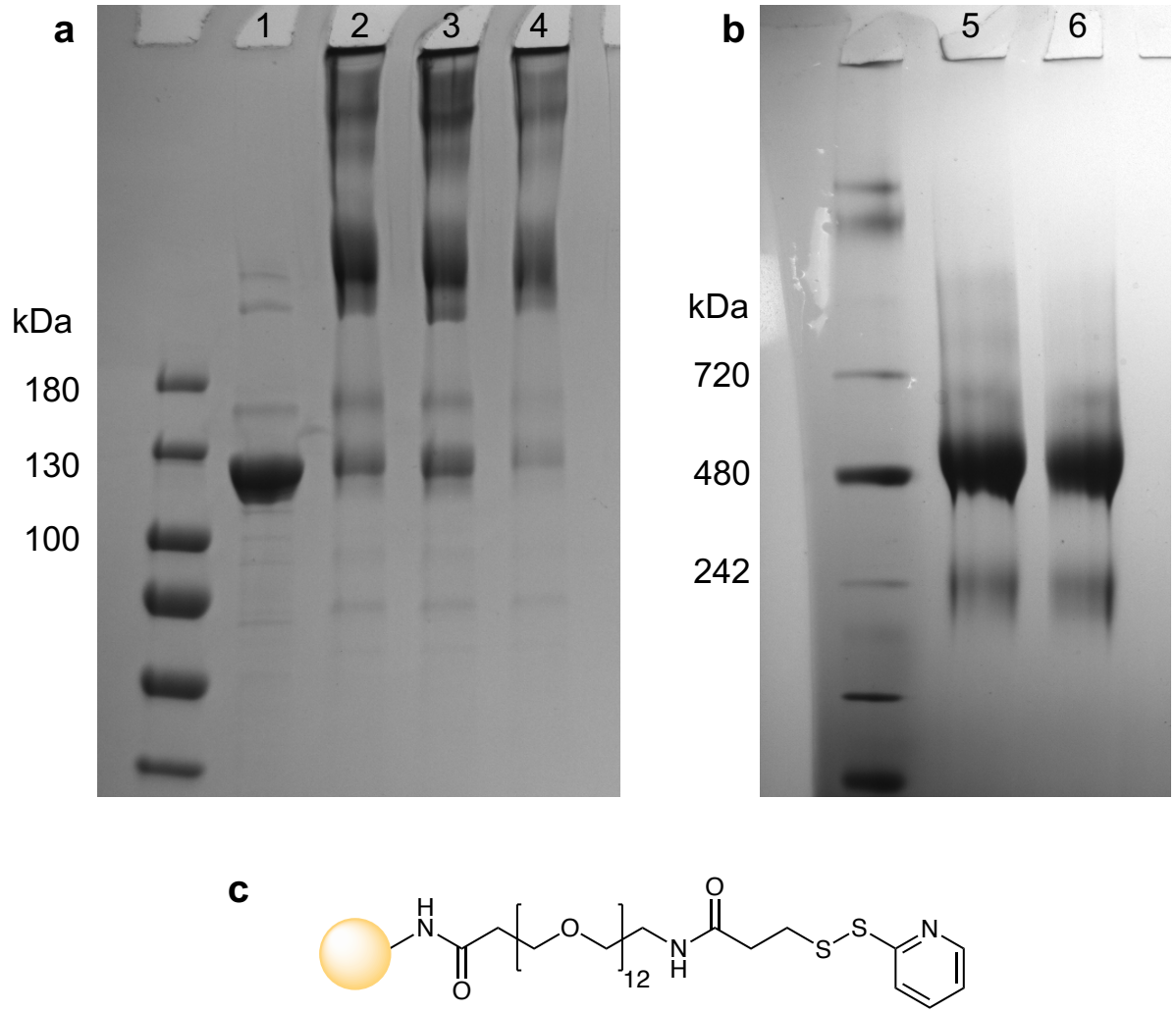



**Figure 4.12 Crosslinking the AbyB3 Tetramer.** a) Non-reducing SDS-PAGE gel of AbyB3 crosslinking by $PEG_{12}$-SPDP. 1) Native AbyB3, with no crosslinking performed. 2) $PEG_{12}$-SPDP conjugated AbyB3, designed to link an amine sidechain with the thiol group on a cysteine residue. 3) 1 and 2 incubated together for 18 hours. 4) $PEG_{12}$-SPDP conjugated AbyB3 is then reduced by DTT, revealing a sulfhydryl moiety attached to AbyB3 *via* an amine sidechain. This is then mixed with unreduced $PEG_{12}$-SPDP-conjugated AbyB3 for 18 h to form an amine-amine linkage. b) Native PAGE gel of crosslinked AbyB3. The reaction from 4 is separated by size using analytical SEC and 5) early peak and 6) late peak fractions identified by UV-vis spectroscopy of the eluent are analysed by native PAGE. c) The $PEG_{12}$-SPDP crosslinker, with the orange sphere representing AbyB3.

Once crosslinked AbyB3 was subsequently visualised by NSEM (Figure 4.13). This revealed a monodisperse, homogenous sample that was taken forward for analysis by cryo-EM.

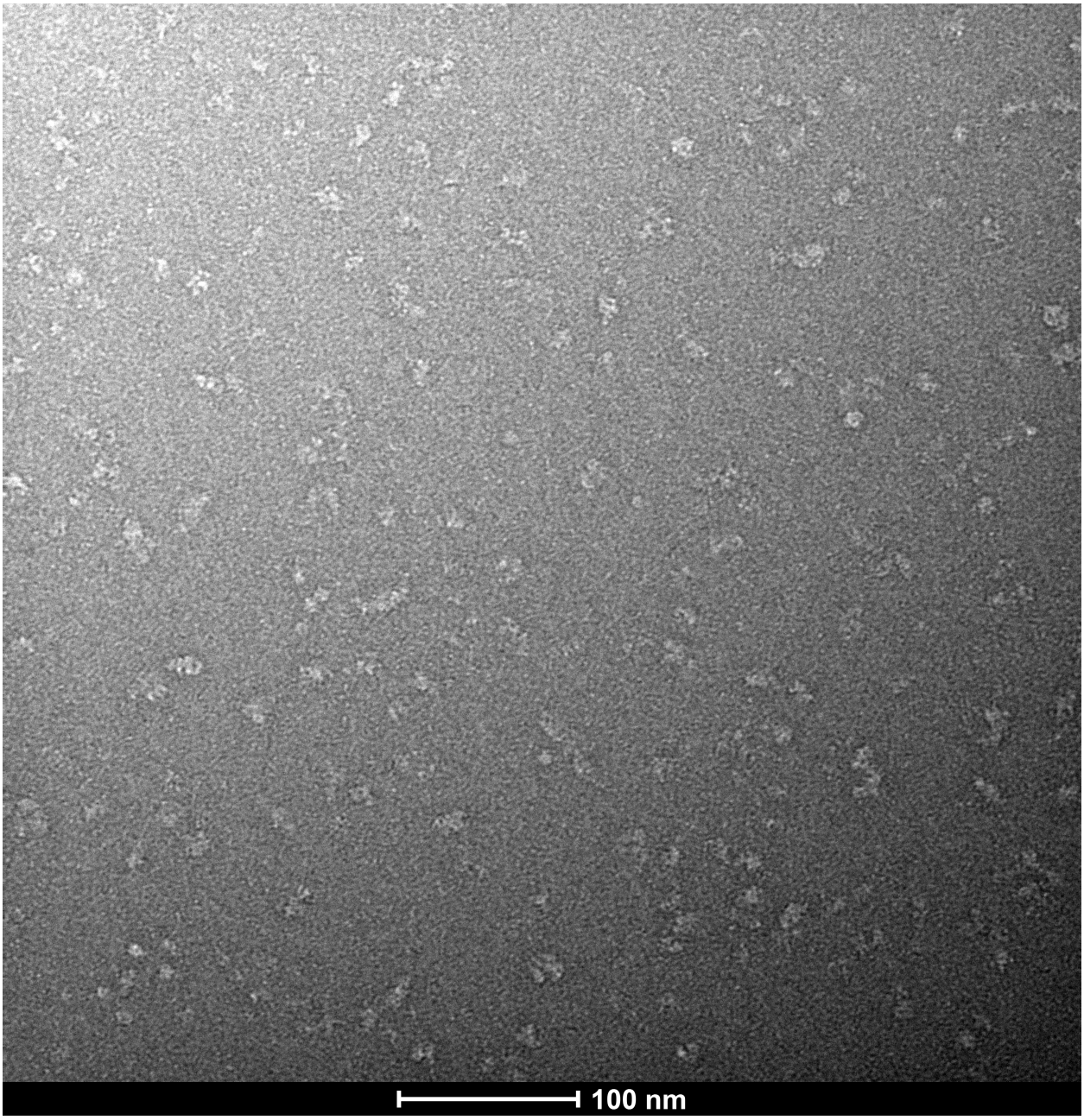



**Figure 4.13 Negative Stain Electron Microscopy of PEG12-SPDP crosslinked AbyB3.** AbyB3 is conjugated by $PEG_{12}$-SPDP, linking two amine residues together. The protein multimers are visualised as white on a dark grey background.

### 4.4.2 Screening AbyB3 Cryo-EM Conditions

A range of conditions were explored in an attempt to analyse native and crosslinked AbyB3 by cryo-EM (Table 4.2). Unfortunately, AbyB3 proved to be a challenging target for study by

cryo-EM. Major obstacles were encountered in attempts to generate monodisperse particles in the layer of ice situated within a quantifoil hole. Both the native and crosslinked proteins adhered to the carbon support around the quantifoil holes, minimising their accumulation within the holes themselves. In an attempt to mitigate this, and to stabilise the protein complexes, R 2/2 grids coated with an ultrathin carbon support were employed. Unfortunately, the small size of AbyB3, even when crosslinked, prevented sufficient contrast from being generated against the support film. In addition, it was feared that the majority of particles were still adhering to the thicker carbon support around the quantifoil holes. To address both of these issues, R 2/2 grids coated in graphene oxide were further treated for use with 1 M magnesium chloride solution. Graphene oxide, because it can be added in a film one atom thick, enables significantly higher contrast for single particle analysis.[299] In order to reduce the hydrophobic nature of standard grids, they are glow discharged, that is, gently ionised at low pressures. When this occurs in air, the carbon supports are given a negative charge.[300] The addition of $MgCl_2$ to graphene oxide coated grids reverses the charge present on the grid, from negative to positive, ideally resulting in an even dispersal of particles. This approach can be used to reduce sample adherence to the negatively charged carbon support. Unfortunately, this did not appear to resolve the issue. Some larger aggregated particles were observed when the solution of crosslinked AbyB3 was incubated on the grid for two minutes, but homogenous monodisperse particles were not observed. Issues with the quality of commercially applied graphene oxide coatings also impeded progress. Some grids were completely uncoated, and others appeared to have had large sections of compromised coating. These issues only came to light after application with protein solution, freezing, clipping and attempted cryo-EM analysis.

Due to these unsuccessful attempts at cryo-EM analysis, a full dataset of negative stain images was acquired for crosslinked AbyB3. 438 micrographs were obtained on an FEI Technai T20 microscope. These were processed using Relion 3.0[294] and particles were picked using a Laplacian of Gaussian difference detector, which is well suited to identifying high contrast negatively stained particles. This yielded a total of ~65,000 particles which were subsequently subjected to multiple rounds of reference free 2D classification. Despite a number of attempts and the use of various strategies, for instance using very large or small boxes to extract the particles and provide them with a localised CTF estimation, or iteratively filtering the particles based on metadata scores determined by Relion, no reliable 2D classes could be generated. This was hypothesised to be because crosslinking AbyB3 with $PEG_{12}$-SPDP did not generate discrete tetrameric complexes, instead linking AbyB3 together in random orientations.

The AbyB3 dimer, which was previously disfavoured due to its lower contrast against the carbon and graphene support films used, was subsequently investigated by cryo-EM using Lacey Carbon grids (Figure 4.14a and b). These grids consist of randomly sized and placed holes within a carbon support stretched between the copper mesh. This delivers a combination of very small and very large holes within the grid; the smallest holes tend to possess extremely thick ice and the largest possess no ice at all. The irregular nature of these grids means that they are not optimised for automated data collection, and it is difficult to obtain datasets large enough to determine high resolution structures by cryo-EM. The automated data collection software is not able to verify its location using the regularly spaced and well-defined quantifoil holes (Figure 4.14c). This results in very small image alignment differences multiplying with every movement of the stage and imaging plane, leading to a highly heterogenous set of micrographs, many of which contain large areas of carbon support or empty holes.

| Type of Grid | Grid Coating | Protein Sample | Instru-ment | Humidity (%) | Incubation *pre*-blot (s) | Blot time (s) | Blot Force | Tempera-ture (°C) | Successful Yes or No | Notes |
|---|---|---|---|---|---|---|---|---|---|---|
| R1.2 / 1.3 | None | AbyB3 & AbyB3 XL | Leica | 100 | 0 | 1 | N/A | 4 | No | No protein visible in quantifoil holes |
| R2 / 2 | None | AbyB3 & AbyB3 XL | Leica & Vitrobot | 80 - 100 | 0 | 1, 2 | 1, 2 | 4, 10 | No | No protein visible in quantifoil holes |
| R2 / 2 | Ultrathin Carbon | AbyB3 XL | Vitrobot | 95 | 0 | 1 | 1 | 4 | No | Background contrast too high |
| R2 / 2 | Graphene Oxide | AbyB3 XL | Leica & Vitrobot | 80 - 100 | 0, 120 | 0.5, 1, 2, 3 | 0.5, 1, 3, 5 | 4, 8, 10 | No | Background contrast too high / no protein in quantifoil holes |
| Lacey Carbon | None | AbyB3 | Vitrobot | 100 | 0 | 0.5, 1 | 1, 3 | 4 | Yes | Protein only visible in thinnest ice far from Carbon edges |

**Table 4.2 Cryo-EM Screening Conditions Investigated.** All grids, apart from those coated in graphene oxide, were glow discharged immediately prior to use at full power (15 – 19 mA) for one minute. 4 μL of protein sample, 0.05 – 0.5 mg/mL, was applied to each grid. Grids were plunge-frozen on either a Vitrobot Mk IV (ThermoScientific) or a Leica EM GP (Leica Microsystems). Grids were screened on a Talos Arctica Cryo-TEM (FEI).

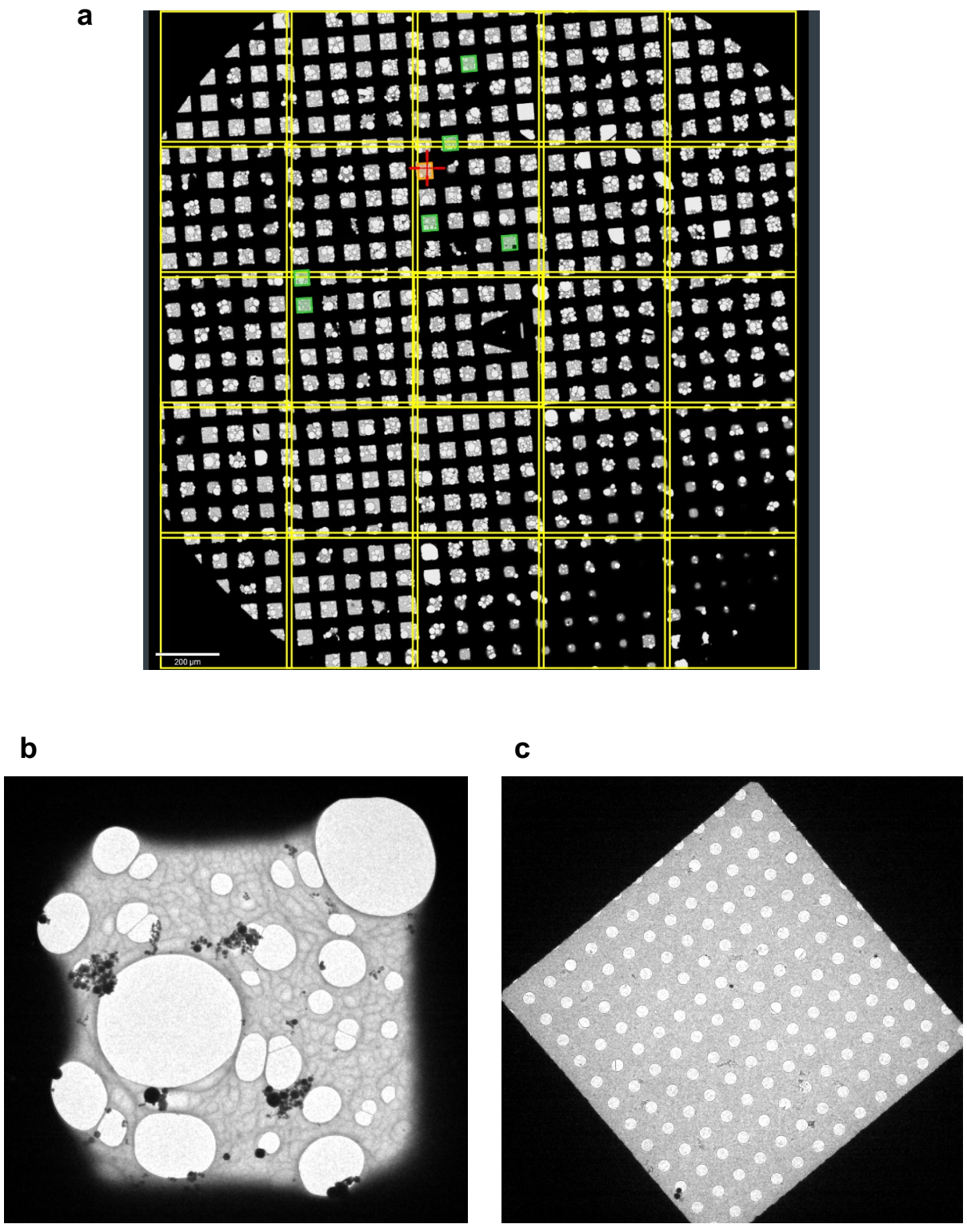



**Figure 4.14 Cryo-EM Atlas and Gridsquare Images.** Images acquired on a Talos Arctica Cryo-TEM (FEI). a) Overview of the entire Lacey Carbon grid used to collect a small dataset. Highlighted in green and red are the seven gridsquares selected for data collection. b) An example Lacey

Carbon gridsquare. The brightest holes are empty and do not contain any ice. Images were taken throughout each selected gridsquare using a computer generated pattern with a 1.5 μm spread between them, deleting selections near carbon support. c) An R 2/2 quantifoil gridsquare.

After monodispersed AbyB3 dimers were observed in some of the initial micrographs obtained on the Lacey Carbon grids, a number of different concentrations of AbyB3, blotting times and blot forces were trialled. These aimed to determine the optimal ice thickness and concentration of AbyB3 needed to observe high-density, yet still monodisperse, particles. The particles appeared to prefer the regions of thinnest ice, as far away from the carbon support as is possible (Figure 4.15). An optimal AbyB3 concentration of 0.12 mg / mL, or 1.12 μM, provides a high density of monodisperse particles within these regions. In order to collect a small dataset of images, a grid was prepared using the conditions outlined in Table 4.3.

| Grid Preparation | |
|---|---|
| Protein / Grid Type | AbyB3 / Lacey Carbon |
| Glow Discharge | 1 min, 19 mA |
| Protein Sample Concentration (mg / mL) | 0.12 |
| Protein Sample Volume (μL) | 4 |
| Plunge Freezer | Vitrobot Mk IV (ThermoScientific) |
| Humidity (%) | 100 |
| Temperature (°C) | 4 |
| *Pre*-Blot Incubation (s) | 0 |
| Blot Time (s) | 1 |
| Blot Force | 1 |

**Table 4.3 Grid Preparation Data for Cryo-EM Microscopy Analysis of AbyB3.**

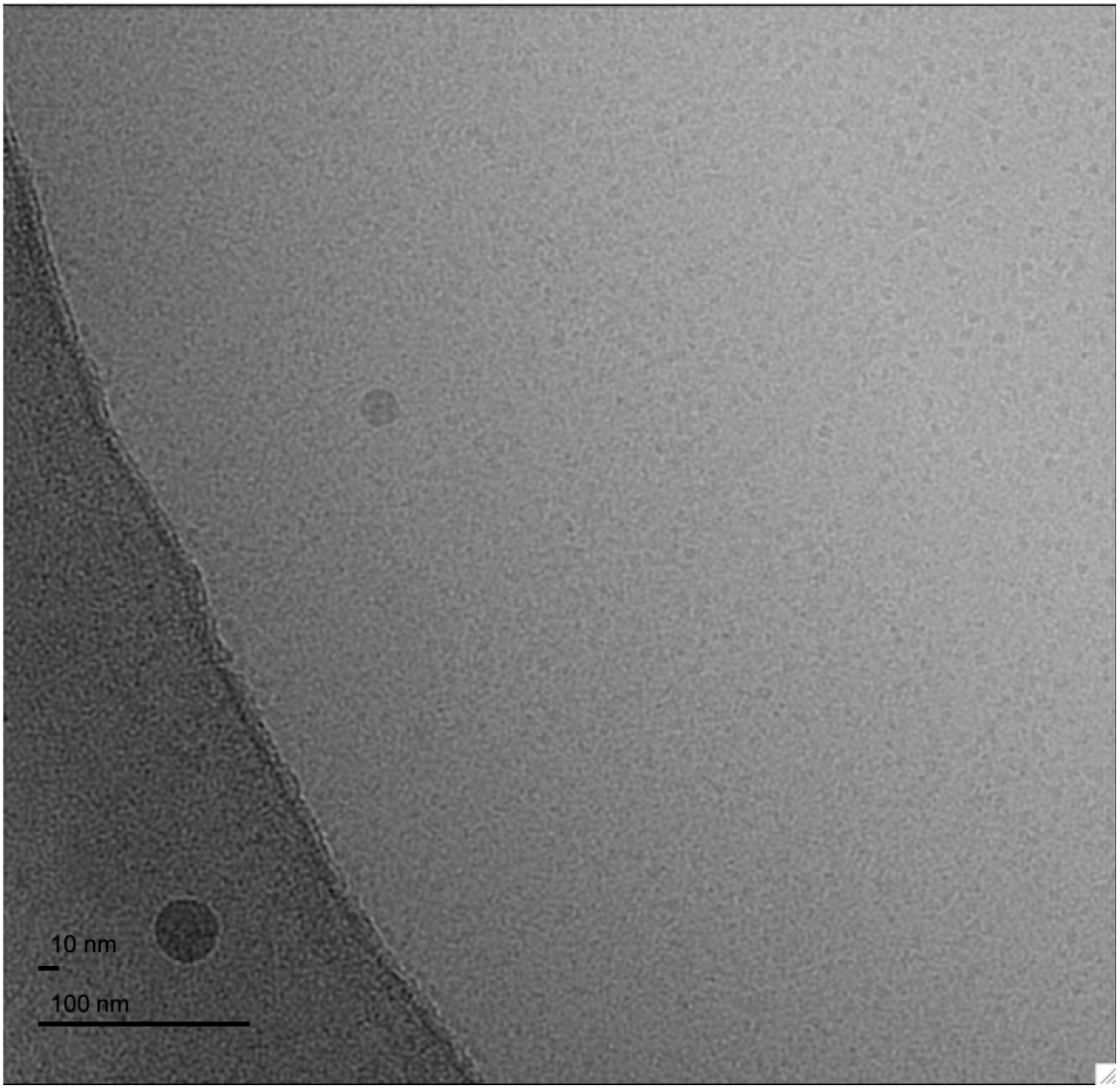
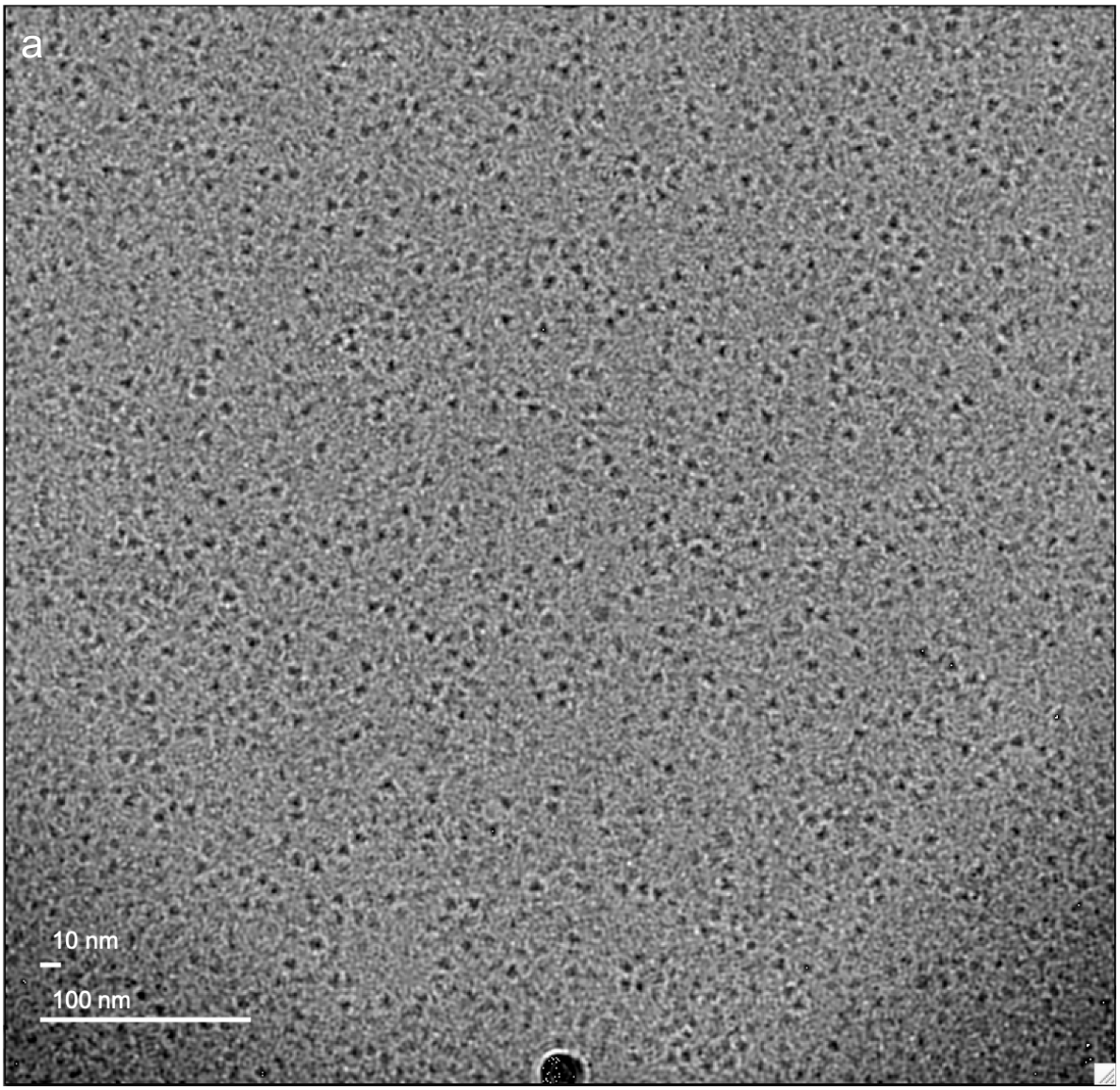



**Figure 4.15 Screening AbyB3 Cryo-EM Grids.** A motion corrected and CTF adjusted micrograph displaying particle preference for thin ice, far away from carbon support. The particles can only be distinguished in the top right hand corner of the micrograph, displaying a small, triangular shape.

### 4.4.3 AbyB3 Dataset Collection

A dataset of 213 micrographs from the prepared Lacy Carbon grid was acquired overnight on an FEI Talos Arctica Cryo-EM microscope (Figure 4.16). The data collection statistics are presented in Table 4.4. Unfortunately, halfway through the overnight run, a camera malfunction resulted in the bottom ~100 nm of the remaining micrographs becoming corrupted and lowering the contrast displayed on the rest of the micrograph. This corrupted section has been

cropped out (Figure 4.16b) and does not impact on downstream processing of particles as extracted particles are normalised with respect to a small surrounding area.



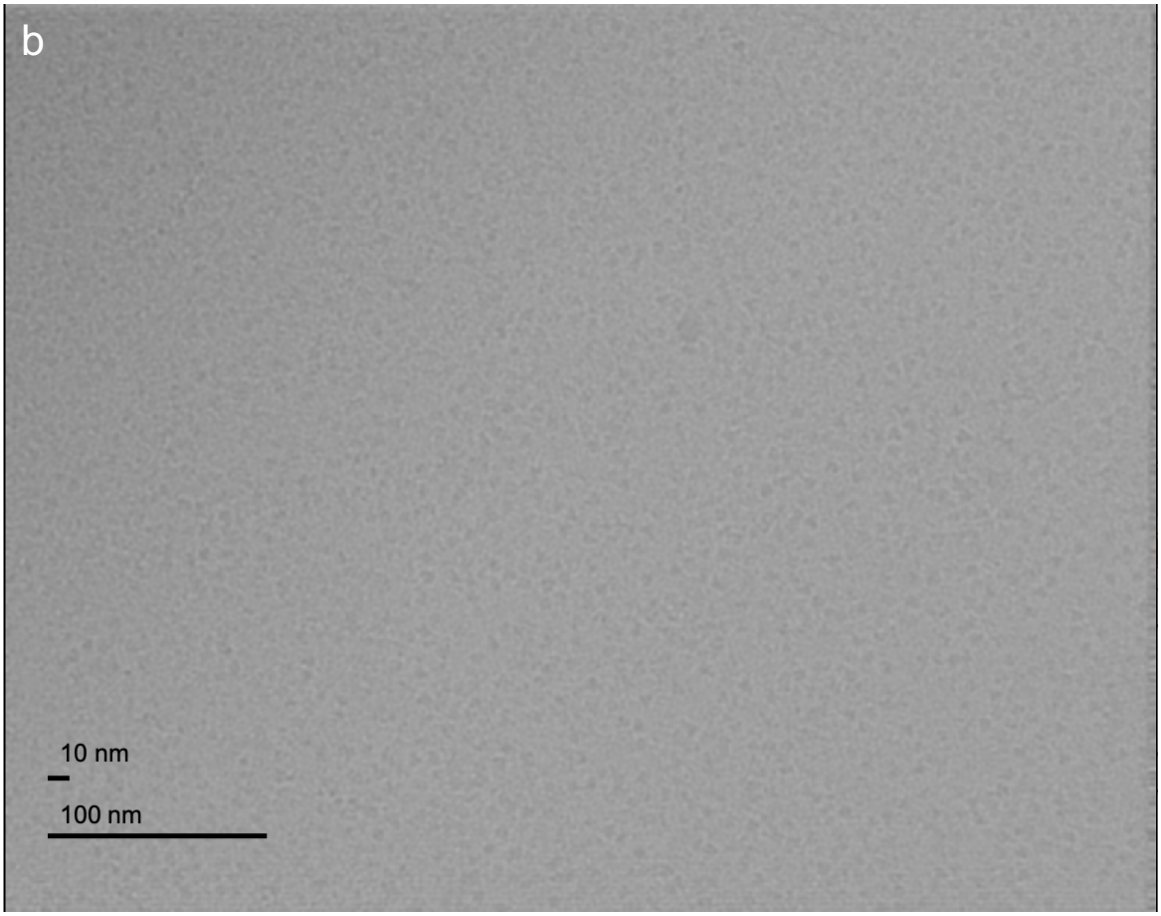



**Figure 4.16 Representative Cryo-EM Micrographs of AbyB3.** Each micrograph was obtained on a Talos Arctica Cryo-TEM (FEI). a) A high contrast micrograph obtained at -2.6 μm defocus. b) An error with the camera resulted in some micrographs displaying as cropped, low-contrast images.

| AbyB3 Data Collection | |
|---|---|
| Hardware | |
| Microscope | Talos Arctica |
| Detector (Mode) | GATAN K2 (Counted) |
| Accelerating Voltage (keV) | 200 |
| Cs (mm) | 2.7 |
| Illumination | |
| Spot Size | 6 |
| C2 intensity (%) | 39.59 |
| C2 Aperture (μm) | 1.8 |
| Dose | |
| Magnification | 130,000 x |
| Pixel Size (Å) | 1.05 |
| Square Pixel ($Å^2$) | 1.1025 |
| Dose Rate (electrons $pixel^{-1}$ $s^{-1}$) | 5.973 |
| Dose Rate (electrons $Å^{-1}$ $s^{-1}$) | 5.377 |
| Exposure Time (s) | 11 |
| Total Dose (electrons $(Å^2)^{-1}$) | 59.147 |
| Number of Fractions | 55 |
| Total Dose per Fraction (electrons $(Å^2)^{-1}$) | 1.075 |
| Data Collection | |
| Objective Aperture (μm) | 70 |
| Defocus Range (μm) | -1, -1.4, -1.8, -2.2, -2.6 |
| Autofocus Frequency (μm) | 6 |
| Delay after Image Shift (s) | 3 |
| Delay after Stage Shift (s) | 6 |
| Energy Slit Width (eV) | 20 |

**Table 4.4 Data Collection Statistics for Cryo-EM Microscopy Analysis of AbyB3.**

### 4.4.4 Processing AbyB3 Cryo-EM Data

The dataset of 213 micrographs was manually sorted, excluding i) those micrographs containing large sections of carbon support; ii) those primarily containing empty holes; and iii) those without discernible AbyB3 particles. This left 38 micrographs, each containing 80 – 714 AbyB3 particles. These particles were picked by hand, as autopicking software is unable to correctly distinguish the small, low contrast, AbyB3 particles in the highly heterogenous Lacey Carbon grids. A dataset of 11,350 particles was extracted, unbinned, using a box size of 231 Å

and a soft circular mask of 100 Å. The dataset was processed using multiple rounds of reference-free 2D classification, with a soft circular mask of 80 Å and a resolution limit of 15 Å. Particles not conforming to the best classes generated with each iteration were discarded. The particles were further sorted according to the Relion metadata ParticleSelectZScore which provides a reliable way to discard particles with the least homology to the class averages generated. Eventually a dataset of 2978 particles was obtained, each of which fit into a 2D class average displaying clear structural properties (Figure 4.17a). Some of these classes did, however, still possess spikes within the solvent region. These spikes are the result of fitting artefacts and sub-classification of these individual classes was performed to remove them (Figure 4.17b). The 2D class averages obtained are derived from a small dataset, and as such, should be treated only as preliminary data.

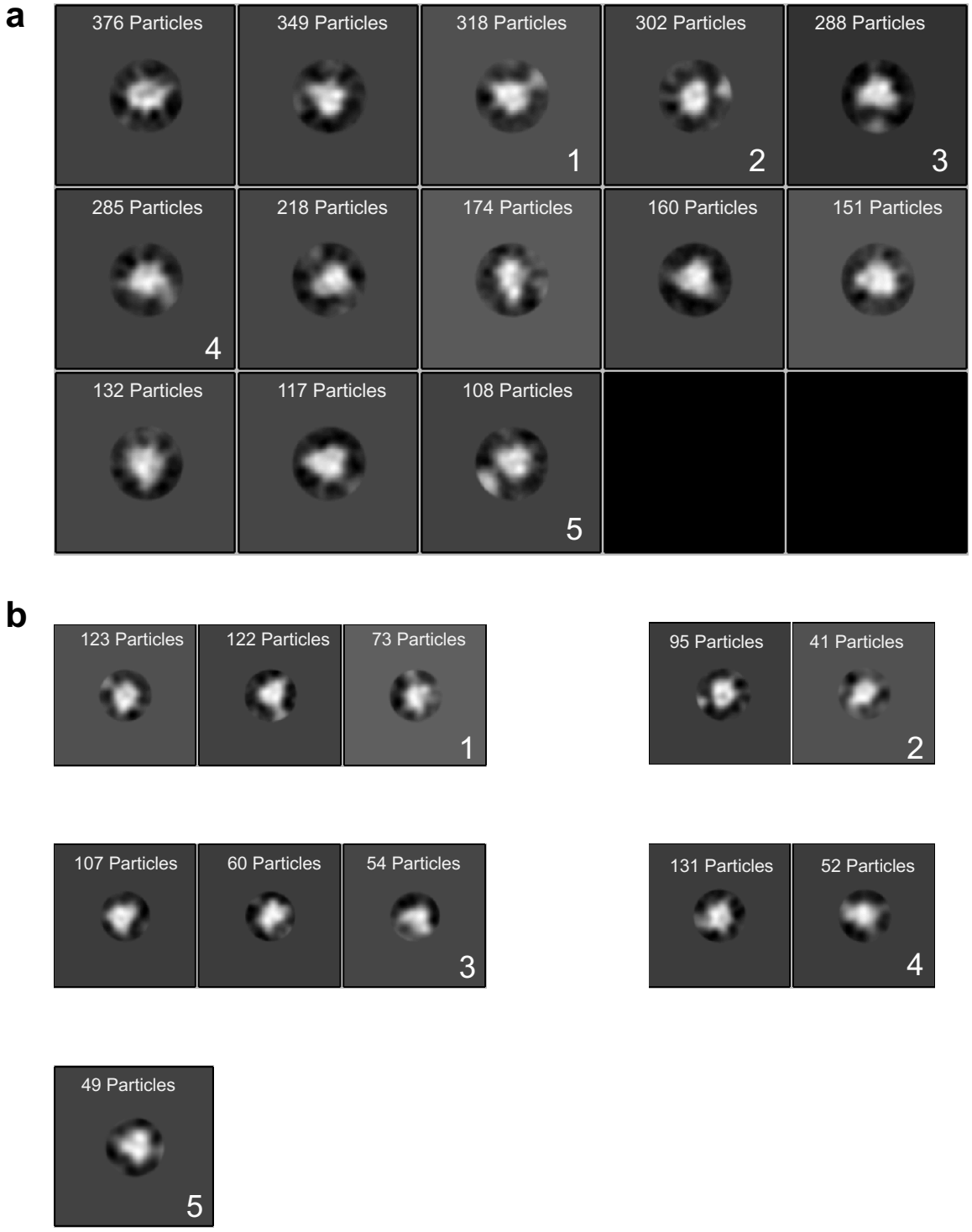



**Figure 4.17 Reference-Free 2D Class Averages of AbyB3.** The micrographs obtained from the Lacey Carbon grids were manually sorted and particles picked by hand. These were then subjected to multiple rounds of reference-free 2D classification until a) a dataset of 2978 particles displaying clear secondary structure was obtained. b) Further sub-classification of 2D class averages displaying peaks within the surrounding solvent (numbered 1 – 5) provided improved class averages. *N.B. These class*

*averages are taken from a small dataset and should not be considered appropriate for the generation of a high-resolution model.*

### 4.4.5 Analysis of AbyB3 Cryo-EM Data

As class averages were obtained with a relatively small numbers of particles (as shown in Figure 4.16), the clear structural features that would be expected from a large publication quality dataset are less evident. Nevertheless, the class averages generated can be compared to those generated during the refinement of the PikAIII model determined by Dutta *et al*.,[141] and the crystal structure of AbyB3ΔACP. The minimal AbyB3 module is expected to be a significantly smaller construct than PikAIII due to a lack of a KR domain. In the PikAIII structure the KR domain is observed at the bottom of the arch shaped structure, creating a large central reaction chamber (Figure 4.3). This would mean that, if AbyB3 adopted a similar structure, it would be expected to primarily consist of the top of the arch observed in PikAIII, potentially forming a triangular shape as observed in the motion corrected micrographs of AbyB3 (Figures 4.15 and 4.16). If, however, AbyB3 adopted a conformation similar to that observed in mammalian FAS[152] and analogous to that observed for AbyB3ΔACP, a more linear shape would be observed in the class averages if only one KS subunit was in view, or an S shaped conformation if two KS subunits were visible.

The low resolution class averages obtained from this initial dataset appear to be in good accordance with those generated by Dutta *et al*. (Figure 4.18). The triangular shape observed in the raw micrographs is representative of the KS and AT domains within PikAIII, when the structure is viewed in a front on arch conformation. Likewise, other 2D class averages match up with those from PikAIII that display a more circular or squashed architecture as a result of the angle of view, and a central cavity can be observed.

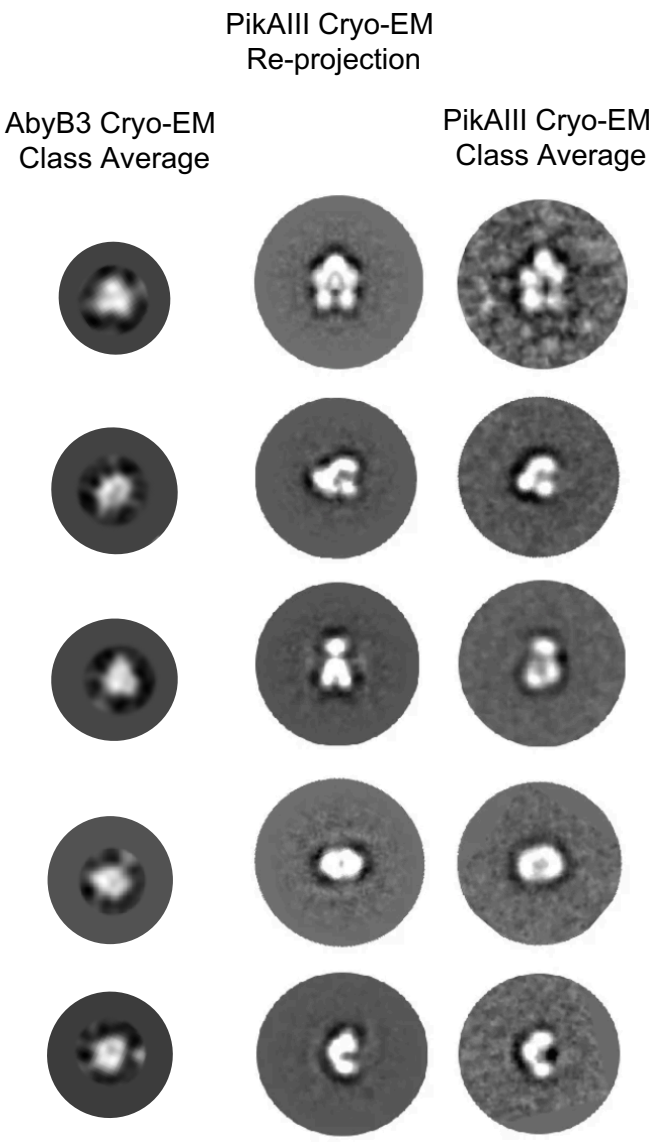



**Figure 4.18 Comparison of AbyB3 Cryo-EM 2D Class Averages with PikAIII.** AbyB3 does not contain a KR domain, which is situated at the bottom of the arch structure displayed by PikAIII. AbyB3 is therefore observed to form a smaller shape, consistent with the top of the PikAIII arch conformation.

The particles adhering to the reference-free 2D class averages obtained (Figure 4.17a) were subsequently used to generate an *ab initio* model of AbyB3 (Figure 4.19). This process typically takes a subset of 400 particles and uses them to generate a very low resolution model of the structure. This is an unbiased process using a sphere the size of the soft circular mask employed when processing the data as a starting point. This initial model immediately indicates that the structure of AbyB3 has significantly diverged from that exhibited by the crystal

structure of AbyB3ΔACP, and is far more representative of the structure exhibited by the KS and AT domains in the cryo-EM structure of PikAIII.

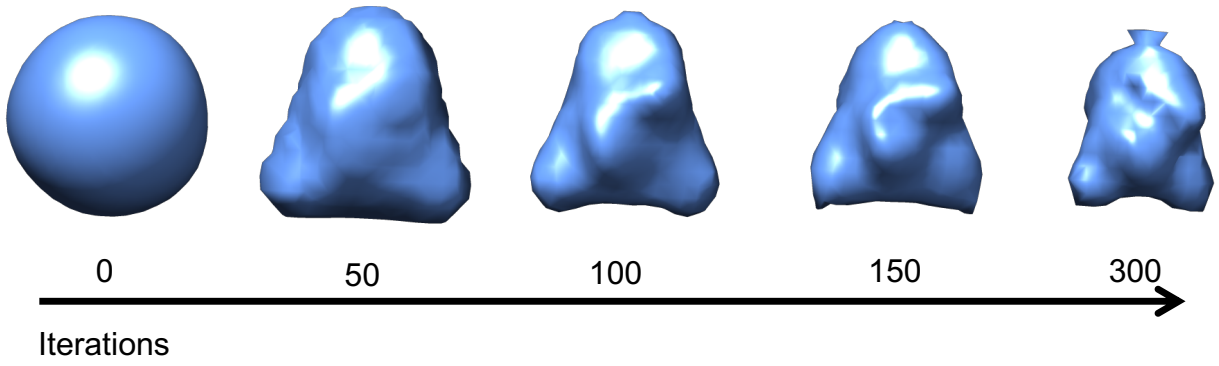



**Figure 4.19 Generating an Ab Initio Model of AbyB3.** The dataset of 2978 particles were used to generate an initial model from an unbiased starting conformation. 300 iterations of refinement yielded a convincing initial model displaying the triangular shape observed in the micrographs.

Although the dataset collected is too small to determine reliable, high resolution, 3D class averages, or to formulate a definitive model, it did yield a number of 3D class averages, under different processing conditions, resembling the 3D class average shown in Figure 4.20a. The initial model and the 3D class averages allow a hypothetical structure of AbyB3 to be proposed, based on that determined by Dutta *et al*. (Figure 4.20b). This shows a closer homology to the 3D class average than the structure of AbyB3ΔACP, although the dataset is not large enough to generate a model that can be reliably docked with the different AbyB3 domains.

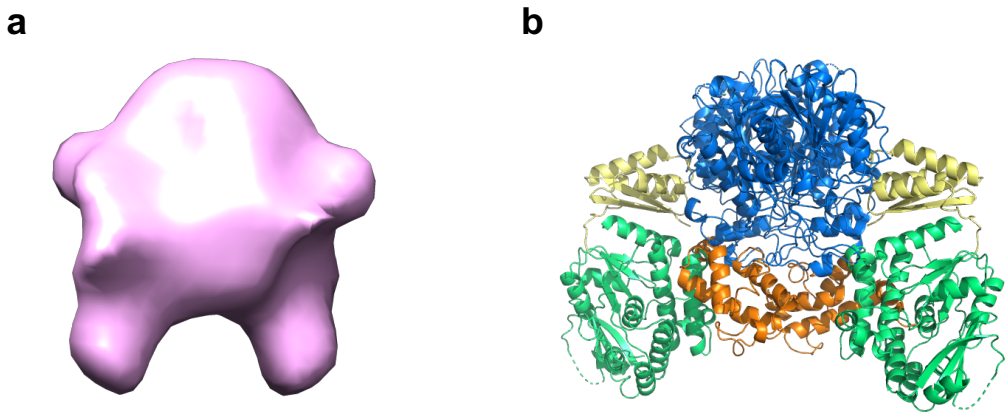



**Figure 4.20 A Hypothetical Model for AbyB3.** a) A 3D class average generated after processing the small dataset shown in Figure 4.17a. A resolution cut-off of 12 Å was imposed and it was generated from 541 particle projections. b) A possible AbyB3 model, constructed in Pymol out of the AbB3ΔACP crystal structure and the structure of the DEBS M2 ACP, using the published structure of PikAIII as a reference.[137] Blue, KS; yellow, KS-AT linker; green, AT; orange, ACP.

## 4.4 Discussion

The abyssomicin C synthase module 7 from *M. maris* has been structurally characterised using both X-ray crystallography and cryo-EM. This structural information is considered critical in providing insight into the mechanism employed by bacterial mPKSs and the molecular details of vectoral biosynthesis. Such studies also provide information about the *inter*- and *intra*-module protein-protein interactions that are known to be critical for mPKS function.

### 4.5.1 Analysis of the Structure of AbyB3ΔACP

The AbyB3ΔACP structure presented herein represents the first high-resolution mPKS didomain structure outside of the DEBS system. Our structure indicates that architectures of these respective mPKS didomains is highly conserved. Although there are differences between the two structures, such as in the placement of salt bridges, or in some of the flexible loop regions, these structures are readily super-imposable. This implies that their mechanism of catalysis is likely to be highly similar, and consistent with that of PikAIII. The AbyB3ΔACP structure sheds light on the structural conservation between bacterial polyketide synthases of diverse species, with *M. maris* and *S. erythraea* sharing only 13.3 % identity by DNA-DNA hybridization.[301] The location of the AT-ACP linker domain that can be discerned is perhaps surprising in that it is identical to the AT-KR linker domain in both DEBS modules. This demonstrates that the association first with the KS-AT linker region, and then with the KS, is likely to be highly specific, with a non-discernible flexible linker then providing the connection to the downstream domain. The relative disorder of the AT-ACP linker domain's interaction with the KS-AT linker could provide some insight as to a possible mechanism by which conformational changes are relayed through the module. The addition of a downstream domain could provide the tension to cause this partial α-helix to dissociate, triggering a rearrangement in the linker region causing the AT domains to bend round, into the arch conformation seen by Dutta *et al*. Specific interactions in this region, influenced by the location of the ACP or the orientation of the KR domain, which Whicher *et al*. showed to be highly variable depending on status of the catalytic cycle, could then result in the movement of the AT-KS linker to obscure the top-most channel to the KS active site, preventing the premature insertion of a polyketide chain and thereby ensuring vectoral synthesis.

### 4.5.2 Analysis of the Preliminary Cryo-EM Structure of AbyB3

Although a full dataset of AbyB3 cryo-EM images is yet to be acquired, the preliminary data presented in this thesis does provide a number of important insights. A significant period of time was needed to investigate a wide range of sample preparation conditions, with the imaging of AbyB3 proving intractable by more conventional EM approaches. A combination of its tendency to adhere to carbon supports, its low contrast, and preference for very thin ice means that most traditional grids proved unsuitable. This presents another possible reason why no cryo-EM structures of polyketide synthase modules have been reported since 2014, as imaging is likely to entail a large resource and time commitment. The successful acquisition of a small dataset does, however, provide significant hope for the future. R 2/1 grids are thought to present the best chance of successful data acquisition using easy-to-process quantifoil grids. Although rarely used for single particle cryo-EM and recommended for the visualisation of an entire cell,[302] the quantifoil holes are spaced more closely together than R 2/2 grids, removing excess carbon support. Larger quantifoil holes are also available, with R 3.5/1 grids an alternative if R 2/1 grids are not suitable. Non-specific crosslinking attempts have been unsuccessful, and future efforts should be concentrated on determining the dimeric structure, despite its smaller size, lower oligomeric number and lower contrast.

The 2D class averages obtained to date are a convincing replication of those obtained by Dutta *et al*. The observation of a triangularly shaped structure in both the 2D class averages and initial model, which are both obtained in a completely unbiased reference-free manner, is a significant departure from the linear or 'S' shaped conformations observed in the structure of AbyB3ΔACP. Although preliminary, the 3D models generated bare good resemblance to what might be expected if an arch shaped conformation was adopted. These models are, however, currently too small to accurately fit in individual domains from the structure of AbyB3ΔACP. This is likely due to a preference for AbyB3 to adopt a face on conformation on the grid, displaying a triangular shape. The small dataset will disfavour any but the most common orientations observed and means that the model generation software is unable to accurately determine the depth of the structure. Hopefully a larger dataset will resolve these issues.

### 4.5.3 Comparison of the Structures of AbyB3 and AbyB3ΔACP

The structure of AbyB3 with and without the ACP appears to differ significantly. The didomain adopts a highly distinctive linear or 'S' conformation that is in good agreement with the prior art. However, the cryo-EM and negative stain analyses preclude this as being the native conformation. AbyB3ΔACP is ~150 Å from the end of one AT domain to the end of the other. Even when crosslinked, distinct particles do not appear to reach this size on a negative stain grid, whilst both negative stain and cryo-EM grids of the non-crosslinked AbyB3 fail to display any particles, even coming close to this size or exhibiting a highly recognisable 'S' configuration. Once processed, the cryo-EM class averages reveal small, mostly triangular, species. Again, none of these class averages resemble the didomain structure and all fit within an 80 Å circular mask. One explanation for the size of these particles could be that the dimer dissociates into monomers upon plunge freezing and vitrification, although this is unlikely. The large KS dimerisation interface, with many hydrogen bonds and two salt bridges is likely to be extremely stable. The close resemblance to the 2D class averages generated by Dutta et al.,[141] as well as the negative stain micrographs, where a method of fixing the particles is employed and is unlikely to precipitate any dissociation events, also indicate that the dimer is observed. An alternative explanation is that the ACP domain, thought to be solely a flexible vector for shuttling the growing polyketide chain between multiple active sites,[151,303] in fact exerts an outsized and unexpected influence on the structure of bacterial mPKSs. The presence of a structurally conserved *post*-AT linker indicates that this region may serve to transmit structural rearrangements from later domains back to the KS and AT domains, revealing and obscuring the upper entrance to the KS active identified by Dutta *et al*. The inability of the same authors to create stable $ACP_4$-$KS_5$-$AT_5$-$KR_5$ constructs without the inclusion of some of the sequence of the ACP from M5 also points to a role in structural integrity for the ACP and the *post*-ACP linker.

Initial studies on AbyB3 and AbyB3ΔACP using SAXS have also shed light onto their respective structures (Figure 4.21). Although these results are still highly preliminary, they do show that, for AbyB3ΔACP, a dimer conformation is exhibited in solution, and that the overall density map generated initially fits well to the 'S' shape observed in the crystal structure. The density map for AbyB3, however, fits far less well to the 'S' shaped dimer it is modelled with, and clearly exhibits a structure with significant differences to that observed for AbyB3ΔACP.

Work is underway to model in each structure with the flexible regions present, as well as with individual domains.

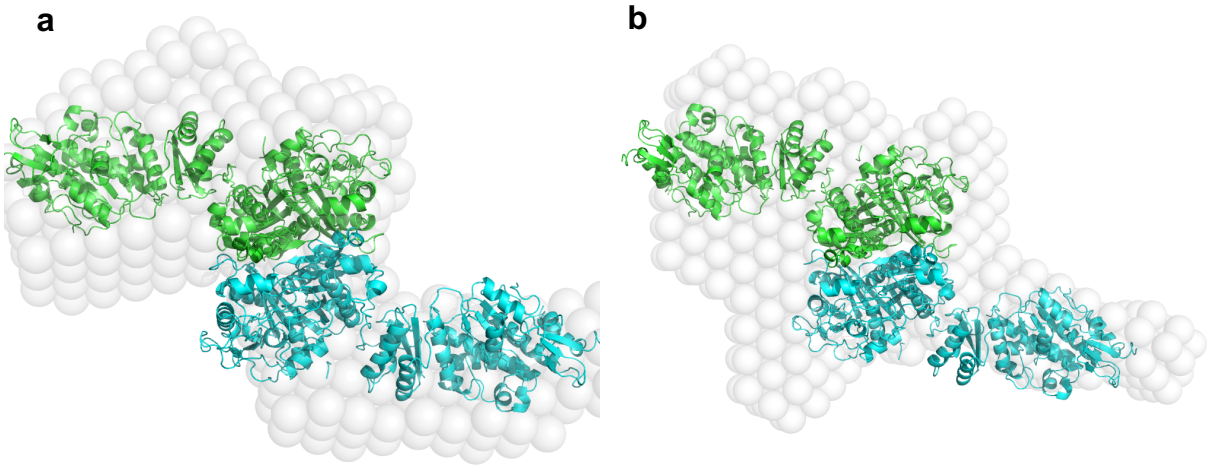



**Figure 4.21 Preliminary SAXS Analysis of AbyB3ΔACP and AbyB3.** Dr. Ash Winter at the University of Bristol has been able to conduct preliminary SAXS analysis of a) AbyB3ΔACP and b) AbyB3 using the initial structure of AbyB3ΔACP for map determination and analysis.

These results show that the structure of a minimal module mPKS should not be predicated on the structure of a didomain. In order to accurately interrogate the structures of modular and multi-modular systems, as well as the highly complex web of structural rearrangements that govern them, they need to be examined as a whole, without any preconceived bias. They also demonstrate, for the first time, the high level of structural homology between mPKS systems from diverse species, and show that there is highly likely to be one overriding structure for mPKS systems. The highly specific protein-protein interactions required for correct vectoral synthesis preclude structural rearrangement and ensure a high level of fidelity between systems.

### 4.5.3 Summary

The work outlined in this chapter constitutes the first comparative structural study between the KS-AT didomain and an intact module of a bacterial polyketide synthase. The cryo-EM

study, although incomplete, is the first structural analysis of a polyketide synthase minimal module, and the second reported of any bacterial PKS module studied by cryo-EM. The two studies both show consistency to the studies that have given rise to two contradictory models for bacterial mPKS structure.[113,141,143,148] Whilst the cryo-EM work remains incomplete, it does provide significant evidence that the addition of the ACP domain precipitates a large structural reorganisation of the module, and casts significant doubt on any model predicting mPKS modules to be structurally homologous to those found in mammalian FASs. This work further exemplifies that, using modern techniques, studies on large, complex biosynthetic machinery originating in little-known and recently discovered species has become possible.

# Chapter 5

# Delineation of the Complete Reaction Cycle of the β-Barrel Diels-Alderase AbyU

## 5.1 Introduction

### 5.1.1 The Diels-Alder Reaction

The DA reaction was originally described by Otto Diels and Kurt Alder in 1928, who identified the two products formed by the reaction of cyclopentadiene with quinone. The development of this reaction, a [4 + 2] cycloaddition, was immediately postulated by Diels and Alder as a potential route to the "synthesis of complex compounds related to or identical with natural products such as terpenes, sesquiterpenes, perhaps even alkaloids". Their work would eventually lead to the award of a Nobel prize in Chemistry in 1950.[304,305] It wasn't until the award of this Nobel prize, however, that the DA reaction began to be used regularly in synthetic chemistry, perhaps due to Diels and Alder's exhortation in their 1928 paper that others were not to use the reaction for the synthesis of NPs. This reaction has since been used extensively in organic synthesis, being used in > 400 syntheses in 2019 alone.[306]

The DA reaction occurs between a 1,3-diene and a dienophile (usually an alkene), to form a cyclohexene (Figure 5.1a). Formal DA reactions proceed concertedly, that is each of the two new carbon-carbon bonds are formed at the same time. This is achieved due to the overlap of the electron orbital fields in the diene and the dienophile. The highest occupied molecular orbital (HOMO) present on the diene overlaps with the lowest unoccupied molecular orbital (LUMO) of the dienophile, allowing electrons to flow between the two (Figure 5.1b & 5.1c).[307] In order for this to happen, electrons are donated by the nucleophilic diene to the electrophilic dienophile. This allows the formation of two carbon-carbon σ bonds at the expense of two π

bonds, and so proceeds in an enthalpically favourable manner.[308] However, in most chemical systems the diene and the dienophile do not have a sufficient electronic interaction for the reaction to proceed, necessitating the presence of electron withdrawing groups on the dienophile, or donating groups on the diene. If these are absent, a catalyst must be used, such as a Lewis acid, which can accept a pair of non-bonding electrons. The Lewis acid (any chemical moiety that can accept a pair of non-bonding electrons) is thereby used to increase the withdrawing nature of the dienophile. These measures strengthen the HOMO-LUMO overlap, allowing the forward reaction to occur. There are also inverse electron demand DA reactions, where the diene becomes the electrophile and the dienophile the nucleophile.[309] The pericyclic reaction remains the same, and likewise has a lack of an energy minimum in the transition state.

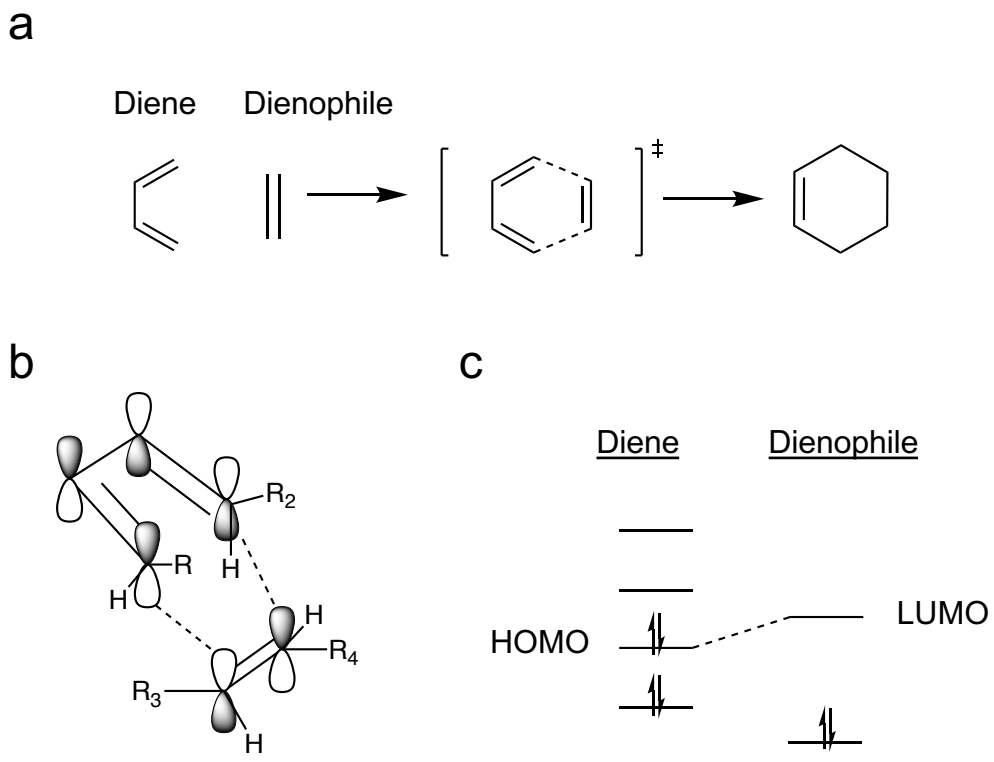



**Figure 5.1 The Diels-Alder reaction.** a) The simplest DA reaction possible, between buta-1,3-diene and ethene to form cyclohexene. The reaction proceeds in a synchronous concerted manner with a single transition state. b) Reaction of the unhybridised P orbitals in the $sp^2$ hybridised ethene and but-1,3-diene. c) Overlap of the HOMO and LUMO electron orbitals between a diene and a dienophile required for DA reaction.

Early progress in the application of DA reactions to synthesis was made using intramolecular reactions, asymmetric variants and hetero-DA reaction,[310,311] as well as in identifying the secondary electronic orbitals that influence the reaction progression.[312,313] More recently, investigations have focussed on the potential applications of the DA reaction in areas such as natural product chemistry and the synthesis of bioactive compounds.[314–317] The DA reaction has also attracted attention as a 'green chemistry' reaction, as there are no by-products produced, and toxic catalysts are typically not needed.[318] The ability of chemists to catalyse this reaction has also been furthered, with the canonical Lewis acid catalysts often used to control the reaction and product stereochemistry recently complemented by novel techniques such as the application of electrostatic fields, shown to drive electron flow from the dienophile to the diene in an inverse system.[304,319] As a result of this extensive theoretical and experimental research, a number of different [4 + 2] cycloaddition transition states have been determined, ranging from synchronous concerted to biradical (Figure 5.2).[307] It has become accepted that for a [4 + 2] cycloaddition reaction to proceed *via* a formal DA mechanism, there must be a concerted bond formation, as well as the formation of a stereospecific product.[320] In this way, a DA reaction must be a pericyclic reaction that proceeds *via* a single transition state.

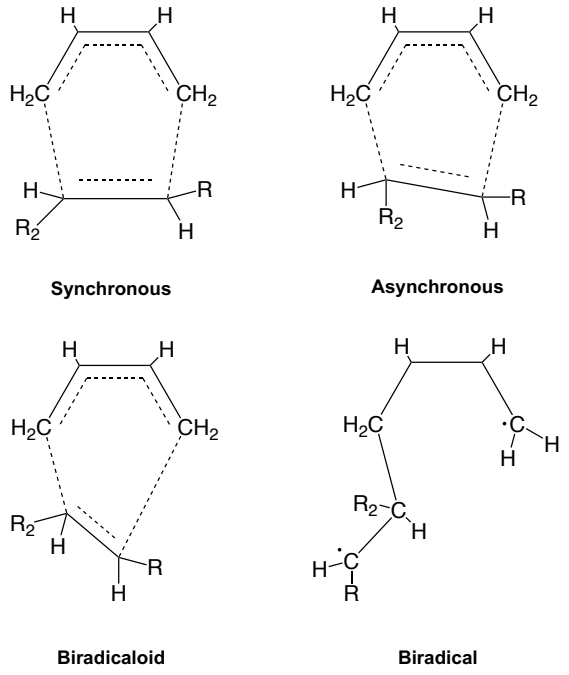



**Figure 5.2 Possible [4 + 2] Cycloaddition Transition States.** Synchronous and asynchronous transition states form bonds in a concerted manner, whereas biradical and biradicaloid transition states do not. Adapted from Klas *et al*.[307]

### 5.1.2 The Diels-Alder Reaction in Industry

Advances in the understanding, control and application of the DA reaction have not, however, translated into the use of this reaction in industrial settings. In 2010 only one marketed drug and two generic drugs in the top 200 approved for use in the USA used the DA reaction in their synthesis.[311,321] This paucity of use is reflected throughout industrial chemical reactions, with the only industry to regularly use the DA reaction in synthesis being the fragrance industry.[322] It is thought that there are two key reasons for the lack of DA adducts in industry. The first is that the substrates are often difficult to work with, and frequently unstable. This not only impacts the ability of chemists to use this reaction at scale, but also comes with a number of safety concerns.[311] Secondly, without using catalysts, it is often difficult to precisely control the stereochemistry of the reaction. For example, in the reaction of methyl acrylate to cyclopentadiene, the products formed are a mixture of *endo* and *exo* species. The addition of a Lewis Acid catalyst, such as aluminium chloride, is sufficient to drive the reaction from an 82:18 to 99:1 *endo*:*exo* ratio.[323] Therefore, if the catalyst used to control the reaction isn't compatible with the rest of the industrial process, or is too expensive or toxic, the DA reaction can yield too great a diversity in product stereochemistry for use in highly stereo-sensitive applications, such as pharmaceutical manufacture.

The simplicity, utility and functional practicality of a reaction that is capable of generating *sp*3 rich cyclic systems without any waste, means that, from the 1970s onwards, research groups have been actively searching for enzymes that have naturally evolved to catalyse DA reactions.[307] Not only would these enzymes be intrinsically valuable in biosynthetic NP pathways, most of which produce cyclised products, but they also might offer a biocatalytic route to producing highly stereoselective DA adducts, without the need to add metal catalysts in bulk. These biocatalysts would likely be far more selective, without potential off-target effects, facilitating the use of the DA reaction more broadly in pharmaceutical and allied applications.

### 5.1.3 Early Enzymatic Diels-Alder Catalysts

Whilst the synthetic DA reaction is dated to 1928, it was a full 67 years before the first potential enzymatic catalyst for this reaction was identified. Oikawa *et al*. successfully showed that a cell free extract from *Alternaria solani* was capable of catalysing a [4 + 2] cycloaddition

reaction in the formation of solanapyrone.[324] Although they were not able to identify the enzyme capable of performing this reaction, or purify it from the extract, they showed that it could precisely control the stereochemistry of the DA adduct formed. They also showed, however, that the same enzyme to catalyse the cycloaddition reaction functioned as an oxidase, preparing the prosolanpyrone precursor for cyclisation. This result was followed up by studies showing that lovastatin nonaketide synthase (a fungal iPKS) was also capable of cyclising a [4 + 2] cycloaddition reaction in the polyketide intermediate it produces, forming a decalin ring system.[325] Macrophomate synthase was hypothesised to catalyse a DA reaction as part of a five-step catalytic process in the formation of Benzoates.

At a similar time to these discoveries a number of other biomolecules were being investigated as putative DA catalysts. These included antibodies[326] and ribozymes,[327] but the characterisation of an enzyme exclusively evolved to catalyse a DA reaction, a standalone DAase, still eluded researchers. Much like the study of PKSs, the identification and characterisation of putative DAases has been almost entirely enabled by gene sequencing technology. This enabled the characterisation of the components BGCs, a key step towards detailed biochemical interrogation of prospective DAases. In some cases, such as that of macrophomate synthase, further studies showed that the enzyme, despite initial encouraging data,[328] was not a DAase, but rather a cyclase.[329] It was shown to create the cyclohexene *via* a Michael addition reaction, followed by an aldol reaction (Figure 5.3a).[330] In other cases, such as in the biosynthesis of the solanapyrones, further studies identified the enzyme Sol5 as the [4 + 2] cyclase necessary to form the bicyclic ring system (Figure 5.3b), however, an additional oxidation reaction was shown to be necessary for the creation of a potent enough dienophile to facilitate cyclisation. It wasn't until the characterisation of SpnF in 2011 as a standalone [4 + 2] cyclase by Kim *et al*., that we could say for certain that enzymes had evolved to promote the formation of cyclohexenes in NP biosynthesis.[331] Heterologous expression of a number of tailoring enzymes in the spinosyn A biosynthetic pathway showed that SpnF was capable, in isolation, of forming a stereoselective cyclohexene (Figure 5.3c).

There were, however, still a number of outstanding questions with regards to DAases. The primary question is whether these enzymes can really be called DAases at all. As studies of macrophomate synthase had shown, and although somewhat semantic, many felt that unless it could be shown that the reaction proceeded in a concerted manner, the enzymes identified should be known as cyclases as opposed to the much sought-after DAases. Counter to this,

from a biosynthetic and industrial point of view, others would argue that if the desired product is produced, the transition states necessary to achieve this are of secondary importance. A further important question is how, if these reactions do proceed *via* a formal DA reaction, is catalysis promoted? Do the enzymes function akin to a Lewis acid catalyst, increasing the withdrawing nature of the dienophile, or do they employ an alternative catalytic mechanism? The length of the search and the high level of interest meant that the 2015 characterisation of three more "elusive and irresistible"[307] putative DAases precipitated the publication of no fewer than 7 reviews into enzyme mediated [4 + 2] cyclisation between 2015 – 2016.[307,320,332–336] These enzymes, VstJ from the versipelostatin NP biosynthetic pathway, and PyrE3 and PyrI4 from the pyrroindomycin biosynthetic pathway were all identified as stand-alone [4 + 2] cyclases.

### 5.1.4 Recent Characterisation of Putative Diels-Alderases

VstJ was successfully purified heterologously in *E. coli* and shown to enable the stereoselective [4 + 2] cycloaddition reaction necessary to create the macrocycle in versipelostatin (Figure 5.3d).[337] The entire BGC was transferred into *Streptomyces albus* for the creation of a Δ*VstJ* knockout strain which produced an accumulation of the cycloaddition precursor. The purification of this precursor and subsequent incubation with VstJ created the formation of the putative DA adduct. Tian *et al*. successfully interrogated some of the final tailoring steps in the formation of the glycosylated and halogenated spirotetramate antibiotic pyrroindomycin B.[338] Within this BGC they identified two stand-alone [4 + 2] cyclases, which act in sequence to create a pentacyclic compound from a linear spirotetramate chain (Figure 5.3e). The first enzyme to act, PyrE3, was identified as an FAD-dependent cyclase. The enzyme was not, however, characterised to work in a typical redox fashion, instead it was determined to use the FAD cofactor purely to stabilise its conformational shape. It was somewhat surprising to find that it could perform catalysis in the absence of NADPH or NADP, especially considering that the enzyme used for homology modelling, OxyS from the oxytetracycline BGC,[339] functions as a typical reductase, hydroxylating at two positions in the presence of NADPH. PyrI4, was, like PyrE3, identified as a [4 + 2] cyclase by homologous expression in *E. coli* and incubation with substrate purified from knockout versions of the host producer, *Streptomyces rugosporus*. Although no initial characterisation of PyrI4 was presented, the same group published the crystal structure of PyrI4 in complex with the product roughly a year later (Figure 5.4).[158] This

revealed a rare β-barrel structure, with the active site within the centre of the barrel. The product was complexed within the barrel and enclosed by an active site capping loop. This loop was shown to be essential for catalysis through mutagenesis experiments. The inside of the barrel was shown to be hydrophobic, with a negatively charged outer surface. The hydrophobicity was hypothesised to facilitate the entry of the substrate into the active site, whilst the positively charged acidic amino acid side chains were postulated to form a more electron withdrawing environment, facilitating the movement of charge from the diene to the dienophile. Whilst together, these two papers from the Liu lab formed the most complete analysis of a hypothetical DAase, it was not until the characterisation of AbyU by Byrne *et al*. in 2016 that any direct evidence of a concerted reaction mechanism was presented.[224]

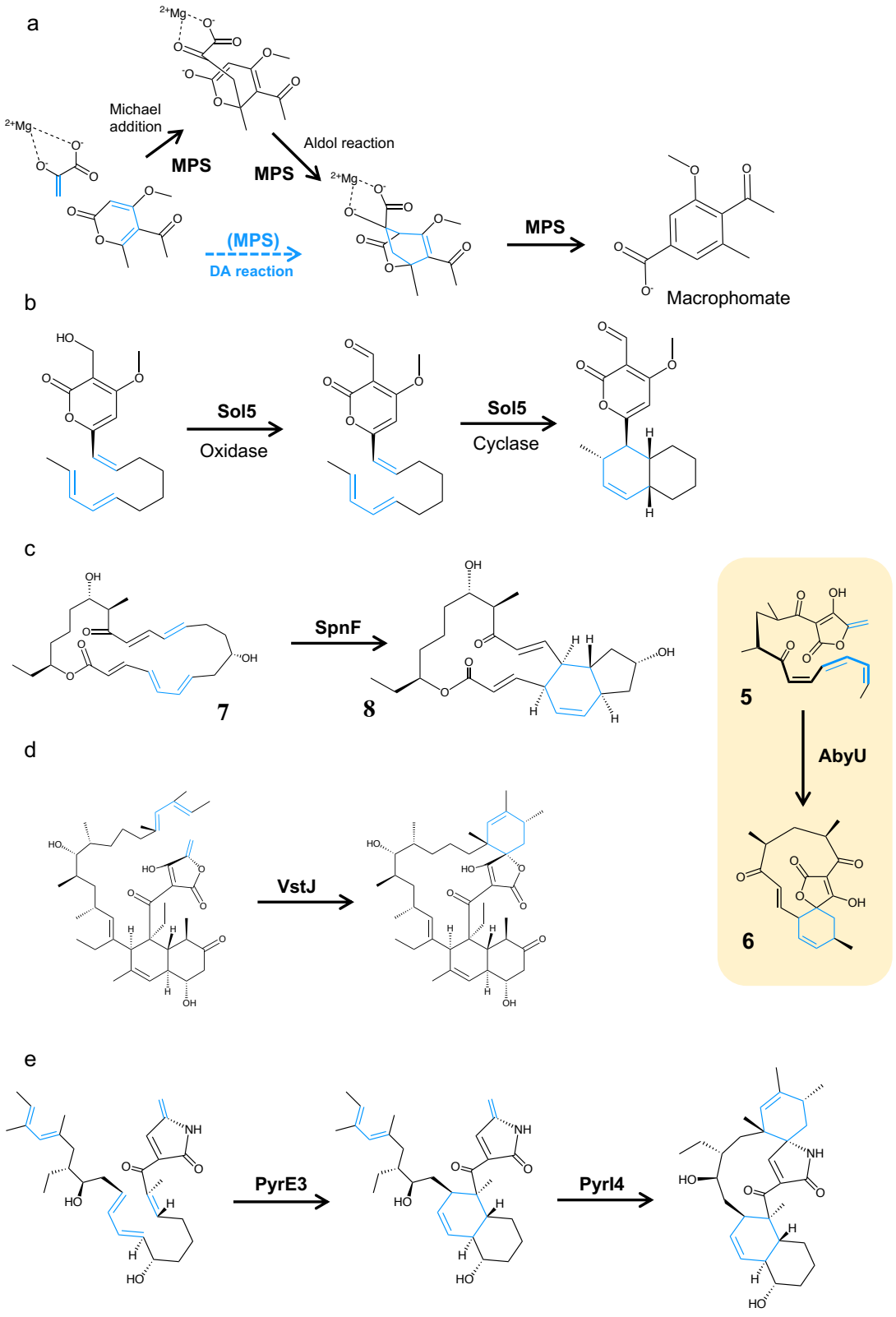



**Figure 5.3 NP biosynthetic reactions hypothesised to be enzyme catalysed DA [4 + 2] cycloaddition.** a) Cycloaddition reaction catalysed by SpnF in the formation of spinosyn A. b) The

cascade of [4 + 2] cycloaddition reactions performed in the biosynthesis of the pyrroindomycins. c) The oxidation and then [4 + 2] cycloaddition steps performed by Sol5 in the biosynthesis of the solanapyrones. d) Cycloaddition reaction performed by VstJ in the versipelostatin biosynthetic pathway. e) Multiple reactions catalysed by macrophomate synthase (MPS) in the biosynthesis of macrophomate. The hypothesised and incorrect DA mechanism is shown in blue. MPS performs a decarboxylation of oxaloacetate (not shown) followed by a Michael-Aldol reaction. The final step in the biosynthesis of macrophomate involves a decarboxylation and simultaneous dehydration. Inset) The DA reaction catalysed by AbyU in the biosynthesis of abyssomicin C.

### 5.1.5 Characterisation of AbyU

AbyU, similarly to other putative DAases, was heterologously expressed in *E. coli* and shown to perform a [4 + 2] cycloaddition reaction *in vitro* by analysis of reactants and products using high-pressure liquid chromatography (HPLC), mass spectrometry and NMR.[224] In contrast, however, this study was performed using synthetic analogues of the natural linear chain **5** (Figure 5.5). This removed the need for large batch fermentation of the slow-growing *M. maris*, but was only feasible due to the relatively simple nature of **9** (Figure 5.5). AbyU, like PyrI4, does not share any sequence identity with non-spirotetronate/tetramate cyclases. It does, however, exhibit the same β-barrel and active site structure as PyrI4 (Figure 5.4). Although Race and co-workers were unable to determine the structure of AbyU in complex with either the substrate or the product, they were able to use computational methods to investigate substrate binding and the mechanism of catalysis. The authors determined 4 possible docked poses of the substrate within the active site. Molecular dynamics showed that this binding occurred in a stable conformation. These poses were then subjected to quantum mechanics/molecular mechanics (QM/MM) simulations to determine the correct pose and the mechanism of the reaction. The authors determined two potentially productive binding conformations, of which one was favoured in both docking and QM/MM simulations (position A, Figure 5.26). The QM/MM simulations of the substrate bound in this conformation showed that the reaction proceeded in a concerted, asynchronous manner, as would be expected for a DA reaction involving a non-symmetrical diene.[307] Higher-level density functional theory calculations were then used to corroborate this data confirming AbyU as the first *bona fide* natural DAase.

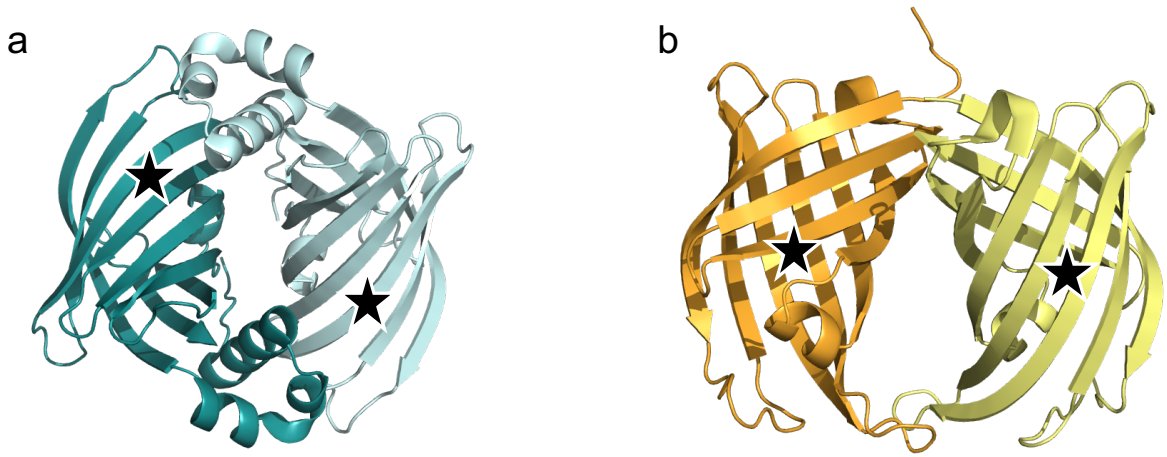



**Figure 5.4 Comparison of the structures of AbyU and PyrI4.** a) PyrI4 and b) AbyU. The active sites at the centre of the β–barrel in each monomer are marked with a star. Both PyrI4 and AbyU exist as a homodimer in solution, although AbyU dimerises in a parallel orientation, with the entrances to the active site adjacent to one another whereas PyrI4 dimerises in an antiparallel orientation with the entrance to each active site at opposite ends of the dimer.

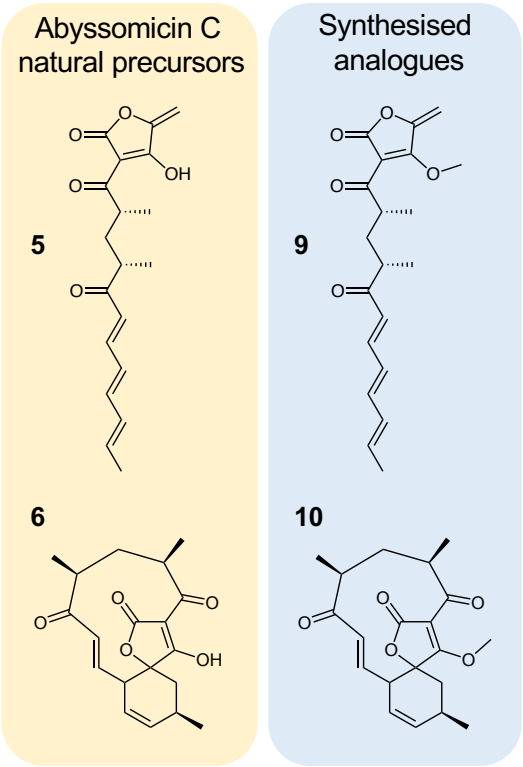



**Figure 5.5 Natural and Non-Natural Substrates Cyclised by AbyU.** Natural abyssomicin C precursors 5 and 6 and their synthesised analogue 9, as well as the biosynthesised analogue 10, created upon incubation with AbyU.

### 5.1.6 Mechanistic Interrogation of SpnF

Following this discovery, a number of further experimental and theoretical studies of SpnF were reported. Patel *et al*. initially investigated the reaction mechanism using QM/MM methods.[189] They hypothesised that the reaction proceeds *via* two distinct intermediates. The first possible intermediate is that of an asynchronous concerted DA reaction, confirming an earlier theoretical study by Hess and Smentek.[340] In contrast, however, Patel *et al*. identified an alternative pathway, believed to be more energetically favourable than a DA reaction. They identified a bis-pericyclic (BPC) reaction intermediate where two cyclic orbital systems stabilise each other and can give rise to two separate cycloaddition products; in this case, a [6 + 4] cycloadduct was formed in addition to a [4 + 2] cycloadduct (Figure 5.6). The [6 + 4] adduct then undergoes a rapid Cope rearrangement to form the [4 + 2] adduct. They concluded that there were a number of potential shallow transition states available in the cycloaddition

reaction for this system, but that the dominant reaction route, according to their simulations, was *via* a [6 + 4] cycloaddition reaction and Cope rearrangement. Medvedev *et al*., in apparent competition with Patel *et al*., then took this theoretical analysis further.[341] They also investigated the possible direct DA and BPC reaction paths, as well as a biradicaloid pathway and an alternative DA reaction pathway. Although Medvedev *et al*. determined that these additional two pathways were not energetically favourable, they were also able to analyse 376 potential BPC and DA transition states. They determined that the BCP transition states were more stable than the DA transitions states, and, in general, possessed shorter *inter*-bond distances. Svitanko and co-workers also identified the BPC transition state analysed by Patel *et al*. as a relative outlier with a comparatively low energy level.[341] In addition, they modelled the reaction of the substrate in water, showing an 83% predominance for a BPC [6 + 4] non-enzymatic cycloaddition reaction over a DA reaction.

The first experimental interrogation of the mechanism of SpnF was undertaken by Jeon *et al*.,[342] who used individually deuterated substrates to monitor the secondary kinetic isotope effect (KIE) during cyclisation. In these complex experiments the authors analysed the rate of formation of each bond in both the enzymatic and non-enzymatic reactions. They were able to deuterate at each α-carbon involved in the putative [4 + 2] cycloaddition reaction (not at all BPC reaction coordinates), and investigate the change in reaction time for the formation of each bond. This was achieved through reaction quenching at specific time points and analysis of the fraction of the reaction that had proceeded during that time. With the heavier deuterium isotopes contributing to longer bond formation times, this enables identification of the rate limiting bond formation step. For a concerted DA reaction, the KIE should be equal for all four deuterated carbon analogues. Liu and co-workers, however, found that, for both the non-enzymatically catalysed and enzymatically catalysed reactions, the reaction proceeds *via* a stepwise mechanism, with the formation of an intermediate.[343] Although this provided no evidence either for or against the existence of a [6 + 4] cycloadduct intermediate, it did reinforce the findings of the theoretical studies, showing that SpnF is a complex cyclase, but not, principally at least, a DAase. All the studies do agree, even if not on the precise mechanism and transition state, that the enzyme is able to function as a catalyst by sterically constricting the substrate in a form close to the transition states.

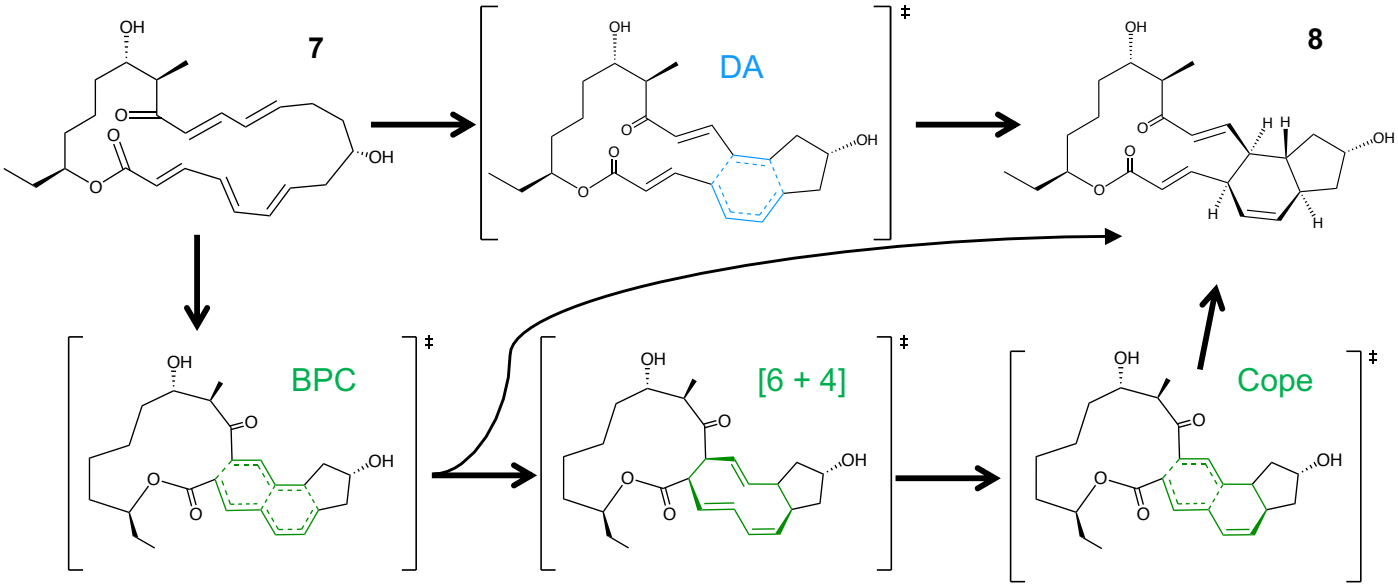



**Figure 5.6 Comparison of Transition States During SpnF-Catalysed Cyclisation.** Patel *et al.*[189] identified multiple possible transition states during SpnF catalysed cycloaddition of **7** to form **8**. The reaction was found to either proceed via a concerted DA reaction, or via a BPC transition state, which could either lead to DA cycloaddition or to [6 + 4] cycloaddition, followed by a Cope rearrangement.

Houk and co-workers were then able to further investigate the theoretical catalysis of **7** to form **8** within the active site of SpnF.[343] This time QM/MM simulations were used to show that SpnF is able to constrain the substrate in a conformation that biases a DA reaction, raising the ratio of DA adduct to [6 + 4] adduct from 1:1 in the gas phase to 11:1 in the enzyme-catalysed reaction. They were also able to theoretically calculate the KIE results obtained by Jeon *et al.*, and find that their results overlapped with good agreement.

### 5.1.7 The Continuing Excitement of Putative Diels-Alderases

This extensive analysis into the mechanism of one enzyme[190,340–343] demonstrates the high level of interest in the mechanism of putative DAases. The 2019 characterisation of two novel [4 + 2] cyclases by Houk, Smith, Sherman, Williams and co-workers in Nature Chemistry further underlines this point.[344] They were able to reconstruct the malbrancheamide biosynthetic

pathway using a combination of synthetic and biosynthetic approaches. They characterised and crystallised the two putative DA homologues MalC and PhqE, revealing a new structural fold for [4 + 2] enzymatic catalysis. MalC was identified as a bifunctional reductase and DAase, although the reaction intermediate has not been thoroughly interrogated, and so the *bona fide* DA nature of the enzyme is still in question. They determined, as is the case with SpnF, PyrI4 and AbyU, that spatial confinement of the substrate leads to catalysis.

### 5.1.8 The Aim of this Work

Despite the high level of interest around the mechanism of putative DAases, there remains significant ambiguity as to the precise mechanisms employed by enzymes that catalyse DA reactions. Even in the case of SpnF, and despite the high level theoretical analysis performed, this enzyme has, to date, not been subjected to a complete biochemical and kinetic characterisation. Although the steady-state kinetic parameters for SpnF, AbyU, VstJ, PyrE3, PyrI4 and MalC have all been reported, these results, often obtained through reaction quenching and HPLC or liquid chromatography – mass spectrometry (LCMS), are by no means a complete study. Although single turnover and burst phase kinetics have been used extensively to determine the kinetic parameters and reaction mechanism for enzymes as diverse as T7 DNA polymerase,[345,346] DNA topoisomerase I,[347] HIV reverse transcriptase,[348] tryptophan hydroxylase[349] and Toluene /o-Xylene monooxygenase,[350] there has been, to date, no extensive kinetic characterisation of a [4 + 2] cyclase or DAase. In order to interrogate the kinetic profile and catalytic mechanism of these systems, this chapter reports a comprehensive single turnover and burst phase kinetic analysis of the *bona fide* β-barrel DAase AbyU. This system represents an ideal case study for the kinetic mechanism of putative DAases due to the ability to heterologously express AbyU in *E. coli* and the accessibility of synthetic diketone **9**.[224] Recent studies have revealed AbyU to be a robust enzyme capable of performing catalysis in a number of different industrially relevant conditions, such as at high concentrations of solvent and high temperatures.[351] This study aims to present a complete kinetic analysis of the system both to provide a new level of insight into [4 + 2] cyclases, particularly those with a β-barrel fold, as well as to provide a rational basis for the mutagenesis and engineering of AbyU, enabling it to become a more attractive industrial biocatalyst. These studies also seek to provide a template for those intending on using a β-barrel scaffolds for the design of *de novo* enzymes.[352,353]

## 5.2 Methods

AbyU was cloned previously according to the method laid out by Byrne *et al*. [224] AbyU was subsequently expressed and purified according to the general method outlined in Chapter 2 Section 3.2.

### 5.2.1 Synthesis of Substrate 9 and Product 10

Synthesis was carried out by Ms. Sbusisiswe Mbatha, University of Bristol according to the protocol outlined by Byrne *et al*.

### 5.2.1 Spectrophotometric Analysis of Substrate 9

A Cary 60 UV-Vis spectrophotometer (Agilent) was used to determine the absorbance profile of both substrate **9** and product **10**, as well as their respective extinction coefficients. **9** and **10**, prepared synthetically as a powder, were individually diluted in 100 % acetonitrile to form a 10 mM stock solution. This stock solution was stored at -70 °C to prevent spontaneous cyclisation, and kept on dry ice when taken out for use. This was further diluted to a 1 mM stock solution on each day of use. The use of more dilute stock solutions was not possible due to increasing rates of spontaneous cyclisation at lower concentrations. To determine the absorbance spectra of each compound, a 100 μM solution was prepared in SF buffer (150 mM NaCl, 20 mM Tris-HCl, 10% acetonitrile, pH 7.5), creating a final concentration of 10 % acetonitrile. 120 μL of each solution was added to a 100 μL cuvette (Hellma) and the absorbance spectra recorded between 200 – 600 nm.

The largest differential between the absorbance profiles of **9** and **10** was determined to be at 325 nm. In order to calculate extinction coefficients at this wavelength, the absorbance profile of a series of concentrations was determined for each compound. The same 100 μL cuvette was used, filled with 120 μL solution to ensure air bubbles did not interfere with the reading. Each substrate concentration was prepared immediately before use, using a 10 or 25 μL Hamilton syringe (Hamilton) to ensure the addition of the correct volume of **9** or **10** diluted in acetonitrile.

### 5.2.2 Dynamic Light Scattering Protocol

Dynamic Light Scattering (DLS) experiments were conducted on a Malvern Zetasizer Nano ZSP (Malvern Panalytical). 110 μL **9** was prepared in a 100 μL absorbance cuvette (Hellma) at a variety of concentrations, diluted in 100 % acetonitrile. A monochromatic light source at 250 nm was used to determine scattering properties of the solution.

### 5.2.3 AbyU Stopped-Flow Experiments

Stopped-flow (SF) experiments were performed using a Hi-Tech Scientific KinetAsyst Stopped-Flow System spectrophotometer (TgK Scientific, Bradford on Avon, UK). Both syringe 1 and 2 were washed with dd$H_2O$ and SF buffer. 10% acetonitrile was maintained in both syringes at all times to ensure solubility of the substrate. The stop-flow was zeroed and references were taken using SF buffer in each syringe for burst phase kinetics. AbyU, in SEC buffer following purification by SEC, was diluted to create a final concentration of 800 μM AbyU and 10 % acetonitrile. This solution of AbyU was present in syringe 1 to obtain references for the single turnover experiments. For the multiple turnover experiments, AbyU was added to syringe 1 at a concentration of 2 μM, diluted in SF buffer. Likewise, substrate **9** was diluted in SF buffer to double the reaction concentration and added to syringe 2. 50 μL from each syringe was injected into the mixing chamber and measurement immediately commenced.

Reactions were observed using a logarithmic timebase over either 1000 s or 10 s. For the 1000 s experiments, the timeframe was split into 15 cycles, each with 200 points, each of which was distributed in a logarithmic manner. This ensured a high number of readings in the first seconds of the experiment. It did not prove tractable to measure each reaction over 1000 s as over this timeframe, **9** can undergo significant background cyclisation to form **10**, rendering the rest of the syringe useless. This was compounded by the slight affinity of substrate **9** for the plastic components in the SF system, rendering it impossible to precisely replicate the concentration of **9** in the reaction chamber over multiple syringes. To measure the uncatalysed rate of substrate cyclisation, the stop flow was zeroed and references were taken using SF buffer in both syringes. Varying concentrations of substrate **9** were added to syringe 2 and measurements were taken using a linear timebase over 10 seconds. All reactions were performed at 22 $^{o}$C.

### 5.2.4 Fitting and Analysis of the Stopped-Flow Transients

Results were fit using Graphpad Prism. Data obtained before 2 ms was discarded. The burst phase raw kinetic data were fit to two exponentials with a sloping baseline (Equation 5). This fit did not, however, reliably determine the angle of the slope observed in the single turnover 10 s experiments due to the shallow nature of the slope and bias introduced by the exponential phases. To circumvent this, the data between 3 s and 10 s were separately fit to a straight line to determine the initial rate of $k_{obs3}$ (Equation 1). Further fitting details are described in the text.

In order to fully analyse the change in the reaction with respect to the concentration of **9**, 3-6 transients at each substrate concentration were taken and the rate and amplitude parameters derived from the fit appropriate for each transient, as is outlined in Chapter 5, Section 3. The background rate of reaction was removed from the steady-state reaction and from the linear portion of $k_{obs3}$. The other exponential phases were too quick for the background rate of reaction to have an appreciable effect. The amplitude of the change in absorbance was converted to concentration using the following measured absorbance coefficient for substrate **9**: $\varepsilon_{325\ nm}$ = 13200 $M^{-1}$ $cm^{-1}$ and the following measured absorbance coefficient for product **10**: $\varepsilon_{325\ nm}$ = 1085 $M^{-1}$ $cm^{-1}$. The rate and amplitude parameters were averaged across all transients from the same concentration of **9**, before plotting against the initial concentration of substrate. Initial concentrations of substrate **9** were calculated from the absorbance at t = 0, to account for un-catalysed substrate turnover, and the tendency of the substrate to have a loose affinity with plastic present in the stop-flow apparatus. Once plotted the data were subsequently fit as is outlined in Chapter 5, Section 3.

In order to calculate the rate of spontaneous cyclization by **2**, the transients were each fit to a straight line (Equation 1). The slope of the line was determined for each substrate concentration and averaged over 3 – 5 repeats.

### 5.2.5 Stopped-Flow Experiments Conducted at 5 °C

Experiments were conducted in a similar manner to the 1000 s single turnover experiments conducted at 22 °C. The KinetAsyst Stopped Flow Spectrometer was additionally equipped

with a T-Pod (also TgK Scientific), which allows individual sets of stopped-flow injections to be incubated at either very high or low temperatures prior to injection. In this case, the reaction mixtures used to create each transient were incubated for 60 s at 5 °C prior to injection.

### 5.2.6 A List of Equations Used

Equation 1:

$$\mathrm{Y} = \mathrm{mX} + \mathrm{C}$$

Equation 2:

$$\mathrm{Y} = \mathrm{Y} = \mathrm{A}_0 + \mathrm{A}_1 \times \exp(-k_1 \times \mathrm{X}) + A_2 \times \exp(-k_2 \times \mathrm{X})$$

Equation 3:

$$\mathrm{Y} = \mathrm{Y} = \mathrm{A}_0 + \mathrm{A}_1 \times \exp(-k_1 \times \mathrm{X}) + A_2 \times \exp(-k_2 \times \mathrm{X}) + A_3 \times \exp(-k_3 \times \mathrm{X})$$

Equation 4:

$$\mathrm{Y} = \mathrm{Y} = \mathrm{A}_0 + \mathrm{A}_1 \times \exp(-k_1 \times \mathrm{X}) + A_2 \times \exp(-k_2 \times \mathrm{X}) + A_3 \times \exp(-k_3 \times \mathrm{X}) + A_4 \times \exp(-k_4 \times \mathrm{X})$$

Equation 5:

$$\mathrm{Y} = \mathrm{A}_0 + \mathrm{A}_1 \times \exp(-k_1 \times \mathrm{X}) + A_2 \times \exp(-k_2 \times \mathrm{X}) + (\mathrm{Slope} \times \mathrm{X})$$

Equation 6:

$$\mathrm{Y} = \mathrm{A}_0 + \mathrm{A}_1 \times \exp(-k_1 \times \mathrm{X}) + (\mathrm{Slope} \times \mathrm{X})$$

Equation 7:

$$\mathrm{Y} = \mathrm{A}_0 + \mathrm{A}_1 \times \exp(-k_1 \times \mathrm{X}) + A_2 \times \exp(-k_2 \times \mathrm{X}) + A_3 \times \exp(-k_3 \times \mathrm{X}) + (\mathrm{Slope} \times \mathrm{X})$$

Equation 8:

$$\mathrm{Y} = \frac{V_{max} \times \mathrm{X}}{K_m \ + \ \mathrm{X} \times (1 + {}^{\mathrm{X}}/_{K_i})}$$

Equation 9:

$$k_{obs} = k_{3\_\mathrm{on}}[\mathrm{S}] + k_{3\_\mathrm{off}}$$

Equation 10:

$$\mathrm{Y} = \frac{A_{lim} \times \mathrm{X}}{K_m + \mathrm{X}}$$

Equation 11:

$$\mathrm{Y} = \frac{A_{lim} \times X^h}{{k_{half}}^h + X^h}$$

Equation 12:

$$k_{obs} = k_{b1\_\mathrm{on}}[\mathrm{S}] + k_{b1\_\mathrm{off}}$$

Equation 13:

6.1 $$k_{obs} = {}^{1}/_{2}\,(p - q)$$

6.2 $$p = k_{b2\_on1} \times [\mathrm{S}] + k_{b2\_off1} + k_{b2\_on2} + k_{b2\_on2}$$

6.3 $$q = \sqrt{p^2 - 4r}$$

6.4 $$r = k_{b2\_on1} \times [S] \times k_{b2\_on2} + k_{b2\_off1} \times k_{b2\_off2} + k_{b2\_on1} \times [S] \times k_{b2\_off2}$$

Equation 14:

$$\mathrm{Y} = \frac{V_{max} \times \mathrm{X}}{K_m + \mathrm{X}}$$

## 5.3 Results: Characterisation of the AbyU Substrate Analogue

In order perform an extensive characterisation of the kinetic parameters relating to the DA reaction catalysed by AbyU, it was first essential to properly characterise the substrate analogue, **9**. In order to monitor the reaction of interest, the cyclisation of substrate **9** *via* a the DA route to product **10** (Figure 5.7a), the spectrophotometric profile of both **9** and **10** were obtained (Figure 5.7b). This revealed significant differences in absorption profiles, and hence the opportunity to measure the progress of the reaction spectrophotometrically. **9** was found to exhibit a peak at ~325 nm, in an absorbance profile between 200 – 400 nm. The absorbance of **10** gave a peak at ~225 nm, in an absorbance profile between 200 – 375 nm.

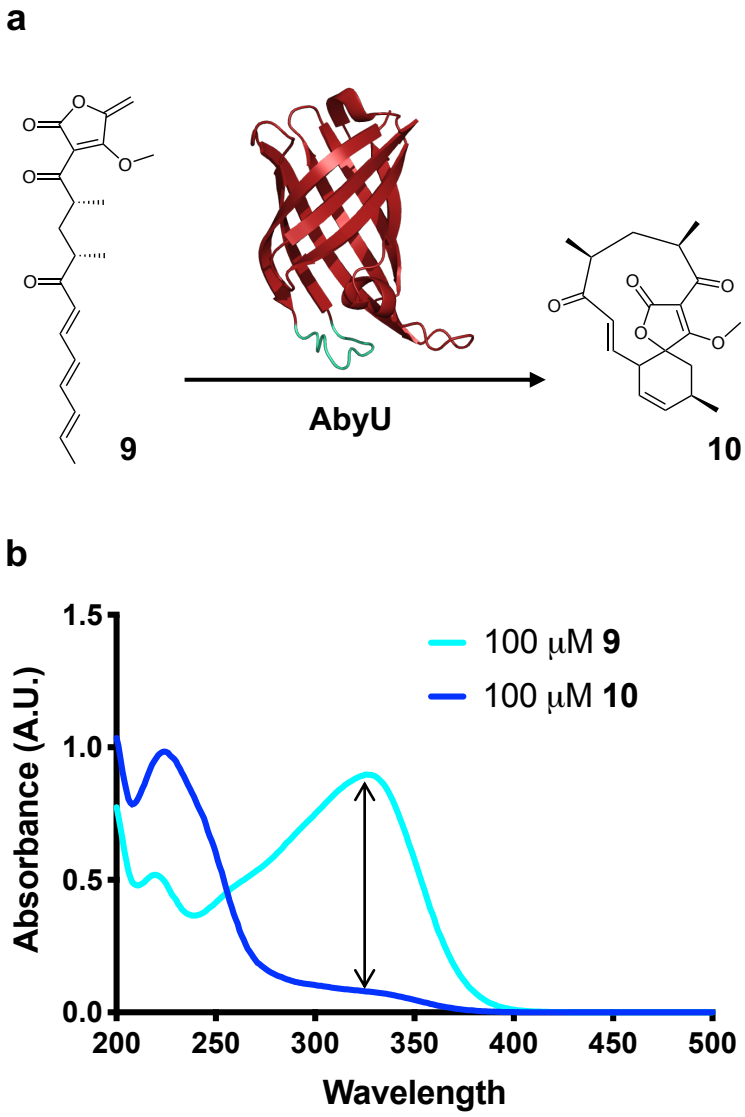



**Figure 5.7 AbyU Catalysed Reaction and Absorbance Spectra.** a) The AbyU catalysed DA reaction to be investigated. The catalytically relevant AbyU monomer is shown with the capping loop highlighted. b) Absorbance spectra of the AbyU substrate analogue **9** and the pre-abyssomicin product analogue **10**. The change in absorbance measured at 325 nm due to cyclisation of **9** is shown by an arrow.

Based on these data, it was considered that the progression of **9** to **10** would be best monitored by following the loss of substrate at 325 nm, as this provides the largest change in absorbance. In order to accurately determine the absorbances of both **9** and **10** at 325 nm, their

extinction coefficients at 325 nm ($\varepsilon_{325\ nm}$) were determined as $\varepsilon_{325\ nm}$ = 13231 $M^{-1}$ $cm^{-1}$ for **9** and $\varepsilon_{325\ nm}$ = 1085 $M^{-1}$ $cm^{-1}$ for **10** (Figure 5.8). Difficulties working with substrate **9** quickly became apparent. Obtaining consistent repeats was difficult and led to extensive testing, with the 6 – 7 most closely related repeats for each concentration shown in Figure 5.8a. These difficulties appeared to be due to the mild affinity of **9** for plastic and the propensity for **9** to cyclise spontaneously at a far greater rate when diluted in buffer. This meant that the establishment of stock solutions of **9** was not feasible, and small, microlitre quantities were added to the cuvette/syringe, as appropriate for the experiments conducted, using a Hamilton syringe. In addition, previous reports from the lab had identified reactions with a higher concentration of substrate as being unable to proceed (unpublished results).

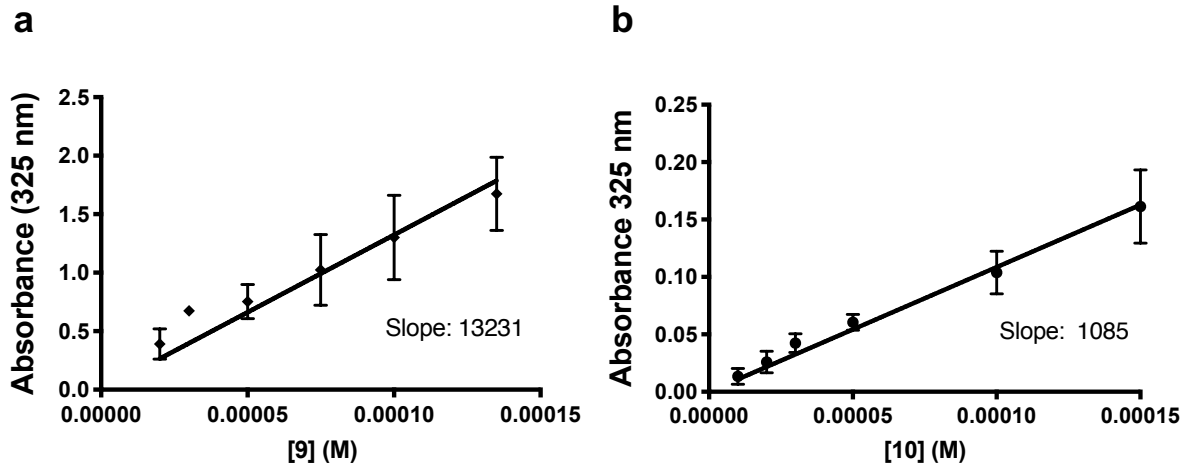



**Figure 5.8 Extinction Coefficients for the AbyU Substrate and Product.** a) Calculation of the extinction coefficient for 9. Each point is the average of 6 - 7 repeats and the error bars are standard deviation from the mean b) Calculation of the extinction coefficient for 10. Each point is an average of 3 repeats and the error bars are standard deviation from the mean.

### 5.3.1 Dynamic Light Scattering Analysis

The propensity of **9** to undergo more rapid spontaneous cyclisation when diluted in buffer was further investigated by DLS. These experiments were motivated by the rationale that **9** chemically resembles a lipid or detergent molecule, with a hydrophobic tail and a hydrophilic head group. It was hence hypothesised that at high concentrations **9** may be able to self-

assemble into vesicle-like structures. DLS experiments were used to establish the presence of higher-order micellular structures that may be formed by **9** in solution (Figure 5.9). These results showed that substrate **9** indeed forms large, regularly sized, structures at concentrations > 500 μM. In the interpretation of DLS results, the identification of a discrete peak is required to ascertain the formation of consistent particles within the solution. The appearance of multiple, randomly dispersed, peaks provides evidence of sample heterogeneity, or the presence of too few particles of a size large enough to scatter light beyond the signal to noise ratio.[354] For concentrations between of 100 – 400 μM **9**, multiple dispersed peaks are observed consistent with sample heterogeneity and the absence of multimeric assemblies. Although the intensity records the presence of larger molecules, the volume distribution implies that these particles are not representative of the entire sample. At concentrations > 500 μM the presence of discrete particles with an average diameter of 160 nm are observed (Figure 5.10). Here the volume, intensity and number peaks are well aligned, implying sample homogeneity. As the concentration of **9** increases up to 30 mM we see an increase in the Z-average size[355] to 329 nm. As might be expected, the size of the multimeric assemblies being formed increases with increasing concentration. For comparison, the globular protein thyroglobulin, with a molecular weight (MW) of 670 kDa, has an average diameter of 8.71 nm, as established by DLS.[354]

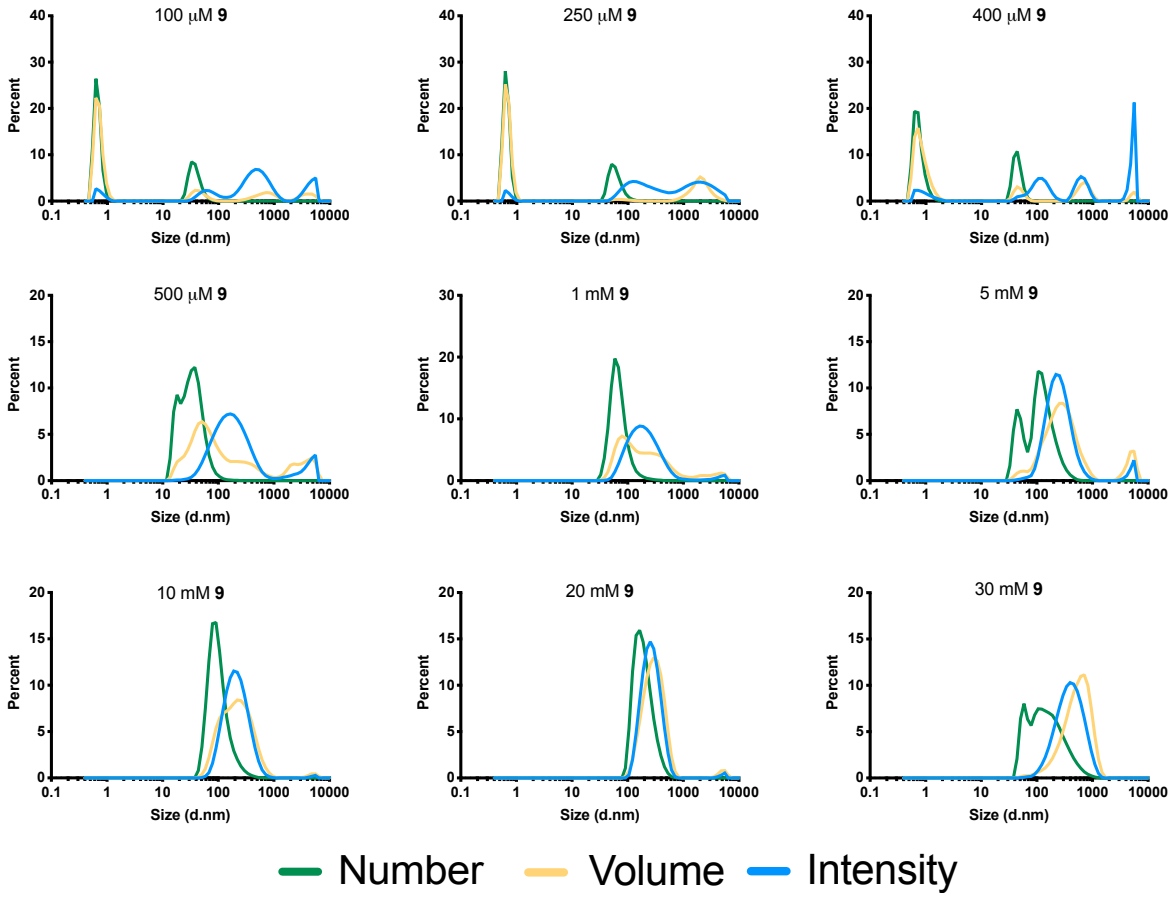



**Figure 5.9 Dynamic Light Scattering of the AbyU Substrate Analogue.** The potential for **9** to form higher order structures was investigated by DLS. Concentrations ranging from 100 μM to 30 mM were investigated. Concentrations of 500 μM and above show evidence for the formation of multimeric structures when measured by Number weighted distribution, Volume weighted distribution and Intensity weighted distribution.

These results clearly show that substrate **9** is capable of forming large, disperse, multimeric constructions in solution. Whether these constructs are the result of polymerisation or the formation of vesicle-like structures is unknown. These structures are significantly larger than AbyU, with a monomer of AbyU likely to measure in the region of 2 nm (carbonic anhydrase, MW = 29 kDa, diameter = 2.37; Aprotinin, MW = 6.5 kDa, diameter = 1.82 nm;[354] AbyU monomer, MW = 17.7 kDa) formations of **9** at 30 mM are around 160 times the size of AbyU. They imply that any work undertaken with substrate concentrations > 400 μM are likely to give

aberrant readings, as AbyU is unlikely to be able to engage substrate molecules which are sequestered within high-order assemblies in solution. There may also be associative effects below this threshold, hence substrate concentrations should be kept as low as possible. It should be noted, however, that these associative effects are likely to protect the substrate from spontaneous turnover when present at high enough concentrations in solution.

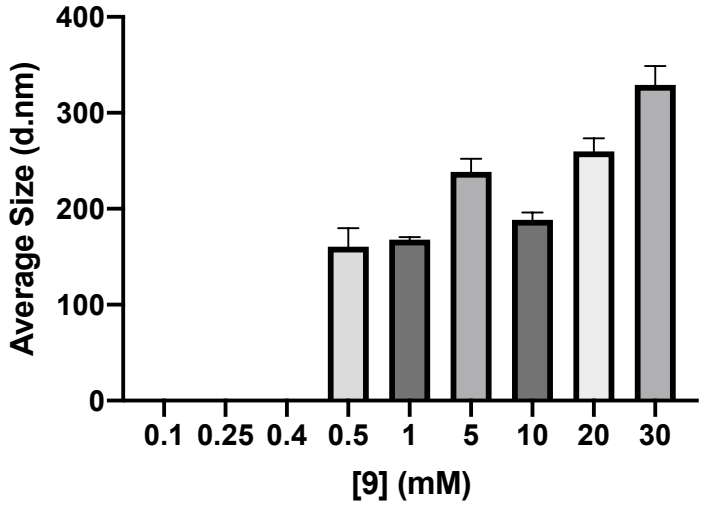



**Figure 5.10 Z-Average Size of AbyU Substrate Analogue Multimers.** The Z-average size is an intensity-weighted measurement for the mean size of particles present and gives a good estimate for the size of particles present, with any noise removed. Results for 0.1 – 0.4 mM were inconclusive and hence are not shown.

### 5.3.2 Spontaneous Cyclisation of the AbyU Substrate Analogue

In complementary studies, the rate of spontaneous cyclisation of **9** in SF buffer was determined (Figure 5.11). These results were obtained by SF to ensure the quality of the data and to mimic the experimental conditions as closely as possible. Spontaneous cyclisation reactions produced a linear decrease in absorbance during each 10 s experiment, with a linear correspondence between the rate of spontaneous turnover and the concentration of substrate,

equating to a non-enzymatically catalysed rate of 0.001255 μM $s^{-1}$. This rate was subtracted from all relevant experiments.

Equation 1:

$$Y = mX + C$$

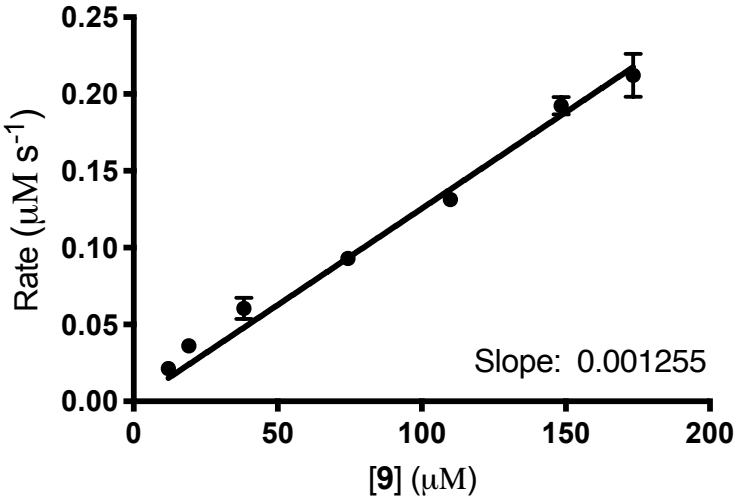



**Figure 5.11 The Rate of Spontaneous Cyclisation of the AbyU Substrate Analogue.** The rate of spontaneous cyclisation of **9** to form **10** was determined in SF buffer. The trend is fit to a line of Equation 1, where m = $(1.26 \times 10^{-3} \pm 2 \times 10^{-5})$, and $C = 0$.

## 5.4 Results: Interrogating the Catalytic Mechanism of AbyU

### 5.4.1 Purification of AbyU

In order to probe the kinetics of the conversion of **9 to 10** by AbyU, the enzyme was expressed in recombinant form in *E. coli* and purified to homogeneity using a combination of nickel affinity chromatography and SEC, as described in Section 2.3.1 (Figure 5.12). Purified protein was routinely recovered in high quantity and at high purity.

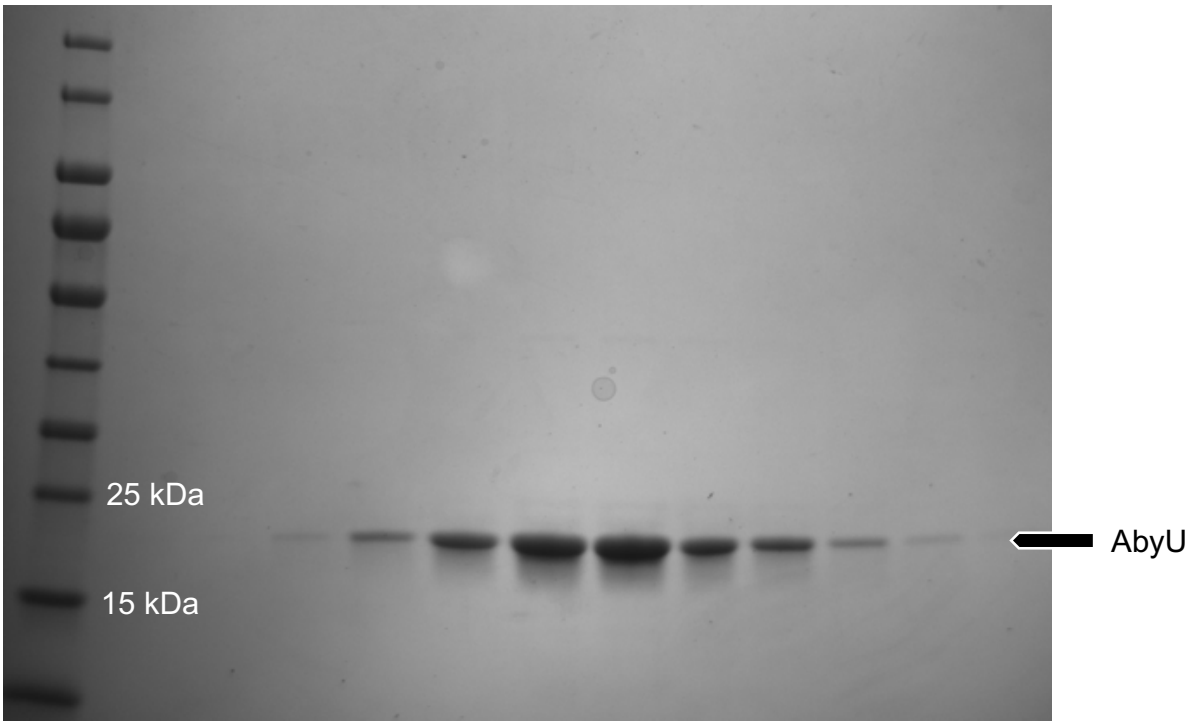



**Figure 5.12 SDS-PAGE Analysis of Purified Recombinant AbyU.** AbyU was purified in SF buffer and estimated to be of > 95 % purity.

### 5.4.2 Single Turnover Experiments Monitoring the AbyU Catalysed Conversion of 9 to 10 over 1000 s

To probe the reaction mechanism of the AbyU catalysed DA reaction (Figure 5.7), two complementary transient experimental methods were used. SF spectroscopy, which has been used extensively to delineate the mechanisms of numerous enzyme classes, and has afforded us a

far greater understanding of enzyme mechanism than steady-state kinetic studies alone.[356] The AbyU catalysed conversion of **9** to **10** was monitored in a SF apparatus both in burst phase, where the concentration of **9** exceeded the concentration of AbyU; and under single turnover conditions, where the concentration of AbyU was greatly in excess of that of **9**. The single turnover experiments were performed to probe the AbyU reaction in a system where each enzyme active site binds and cyclises a maximum of one molecule of substrate. In order to ensure AbyU was significantly in excess of the substrate, a final concentration of 400 μM enzyme was used. Reaction progress was monitored spectrophotometrically at 325 nm for 1000 s (Figure 5.13). Resulting absorbance traces indicated the presence of multiple reaction transients, indicative of an unanticipatedly more complex reaction cycle than was previously predicted.[24] The reaction progresses rapidly during the first 0.5 s, proceeding more slowly beyond this point for the entirety of the 1000 s measurement period. This is characterised by the observation of a shallow exponential curve from 1 s onwards.

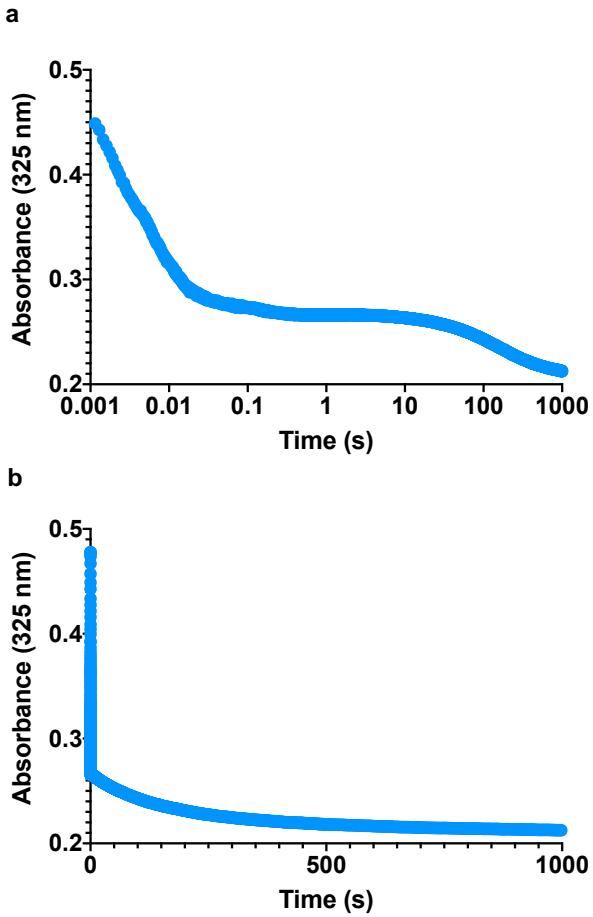



**Figure 5.13 Transients of the reaction of AbyU with 25 μM substrate 9 over 1000 s.** The reaction was monitored spectrophotometrically at 325 nm. Transients are presented against absorbance as both the product and substrate absorb at 325 nm. a) A transient of the reaction between 400 μM AbyU and 25 μM **9** displayed on a logarithmic scale. b) The same transient displayed on a linear scale.

In order to fully characterise the observed reaction transients, kinetic data were fit to three different exponential equations and the quality of data fitting assessed (Figure 5.14). A two exponential fit (Equation 2, Figure 15a), poorly satisfied the initial 0.1 s of the reaction, and a significant non-linear trend can be observed in the residual function. A three exponential fit (Equation 6, Figure 15b) is observed to fit the transient much more accurately, with the adjusted $R^2$ value increasing towards 1. A four exponential fit (Equation 7, Figure 15c) was observed to resolve a slight mis-fitting at the end of the curve, but as this mis-fit would be likely resolved in a 3 exponential fit with additional data points, and because there was no overall significant increase in the fitting statistics, it was decided that the reaction was best fit to an equation with three exponential terms. There were, however, a number of practical difficulties associated with

measuring the reaction over 1000 s. Over this timeframe, **9** is able to undergo significant spontaneous cyclisation to **10**, whilst held at temperature in the syringe of the stop-flow apparatus. This means that subsequent runs from the same syringe were not comparable as repeats. The issues obtaining precise repeats of substrate concentration (as discussed previously) meant that this was not a viable method to investigate the reaction in relation to **9**.

Equation 2:

$$\mathrm{Y} = \mathrm{Y} = \mathrm{A}_0 + \mathrm{A}_1 \times \exp(-k_1 \times \mathrm{X}) + A_2 \times \exp(-k_2 \times \mathrm{X})$$

Equation 3:

$$\mathrm{Y} = \mathrm{Y} = \mathrm{A}_0 + \mathrm{A}_1 \times \exp(-k_1 \times \mathrm{X}) + A_2 \times \exp(-k_2 \times \mathrm{X}) + A_3 \times \exp(-k_3 \times \mathrm{X})$$

Equation 4:

$$\mathrm{Y} = \mathrm{Y} = \mathrm{A}_0 + \mathrm{A}_1 \times \exp(-k_1 \times \mathrm{X}) + A_2 \times \exp(-k_2 \times \mathrm{X}) + A_3 \times \exp(-k_3 \times \mathrm{X}) + A_4 \times \exp(-k_4 \times \mathrm{X})$$

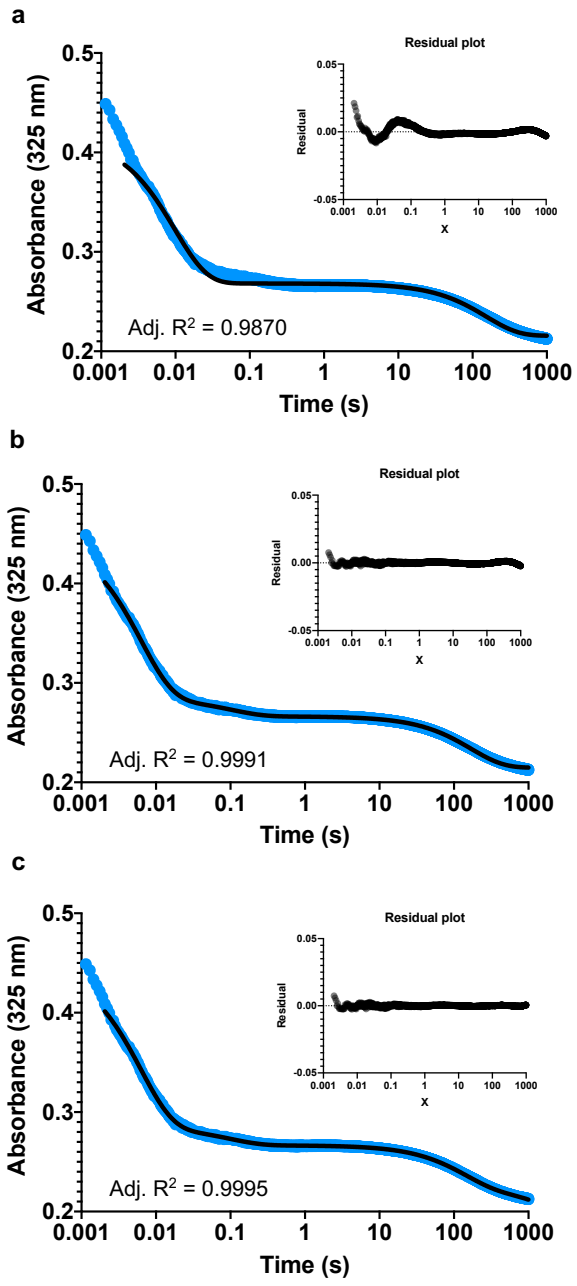



**Figure 5.14 Fitting analysis for the reaction of 25 µM Substrate 9 with 400 µM AbyU over 1000 s.** The residual plots are inset and the adjusted $R^2$ value given below the transients. a) The reaction transient fitted to a biphasic equation (Equation 2). b) The reaction transient fitted to a triphasic equation (Equation 6). c) The reaction transient fitted to an equation with four exponential terms (Equation 7).

### 5.4.3 Single Turnover Studies of the AbyU Catalysed Conversion of 9 to 10 Over 10 s

As the majority of the AbyU reaction occurs within 0.5 s, it was decided that the reaction should be monitored over 10 s to obtain an understanding of how the AbyU catalysed conversion of **9** to **10** is influenced by varying substrate concentration (Figure 5.15).

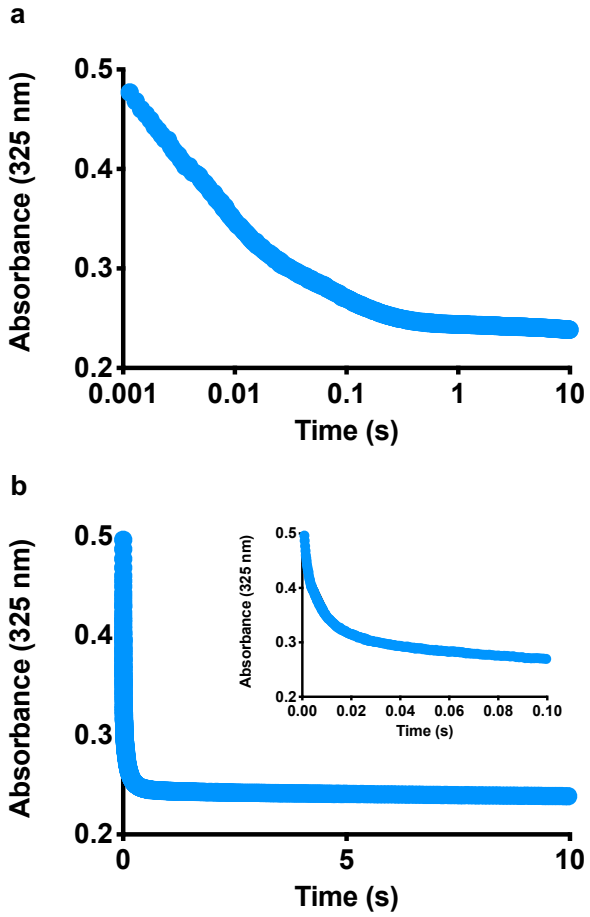



**Figure 5.15 Transients for the reaction of 30 μM substrate 9 with 400 μM AbyU over 10 s.** The reaction was measured spectrophotometrically at 325 nm for 10 seconds. a) The transient displayed on a logarithmic scale. b) The same transient displayed on a linear scale, with the first 0.1 s shown over a linear scale inset.

Reaction transients observed over this shorter time period were found to be in good agreement with those identified in longer duration experiments. However, the final exponential phase observed appears linear over this shorter timeframe. Data fitting analysis (Figure 5.16)

was performed using equations incorporating an additional term for a straight line, to account for the initial long, slow exponential phase. On this occasion an equation consisting of two exponential terms and a straight line (Equation 5) was found to best fit the experimental data, with a single exponential plus slope (Equation 6) introducing significant bias into the calculated residuals. No significant improvement was observed in the fitting data following the addition of a third exponential term (Equation 7). The three exponential phases observed over the longer time frame indicate three separate reaction conformations are formed during the AbyU catalysed conversion of **9** to **10**. These conformations are not observed in the spontaneous reaction, implying that they are specific to the enzyme catalysed reaction.

Equation 5:

$$\mathrm{Y} = \mathrm{A}_0 + \mathrm{A}_1 \times \exp(-k_1 \times \mathrm{X}) + A_2 \times \exp(-k_2 \times \mathrm{X}) + (\mathrm{Slope} \times \mathrm{X})$$

Equation 6:

$$\mathrm{Y} = \mathrm{A}_0 + \mathrm{A}_1 \times \exp(-k_1 \times \mathrm{X}) + (\mathrm{Slope} \times \mathrm{X})$$

Equation 7:

$$\mathrm{Y} = \mathrm{A}_0 + \mathrm{A}_1 \times \exp(-k_1 \times \mathrm{X}) + A_2 \times \exp(-k_2 \times \mathrm{X}) + A_3 \times \exp(-k_3 \times \mathrm{X}) + (\mathrm{Slope} \times \mathrm{X})$$

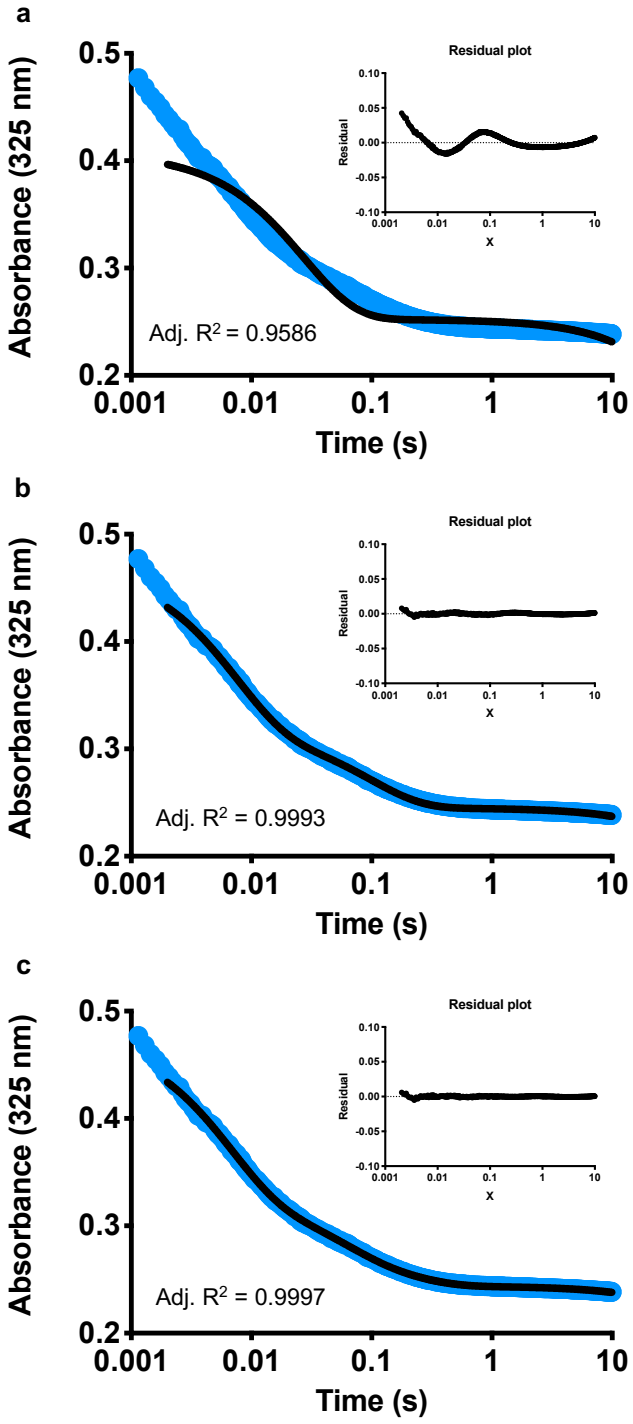



**Figure 5.16 Fitting Analysis for the Reaction of 30 μM Substrate 9 with 400 μM AbyU Over 10 s.** The residual plots are inset and the adjusted R2 values are given below the transients. a) The reaction transient fitted to Equation 6. b) The reaction transient fitted to Equation 5. c) The reaction transient fitted to Equation 7.

### 5.4.4 Interpretation of Single Turnover Reaction Transients

The complex multiphasic transients observed during the single turnover experiments allow determination of three distinct rates of reaction, $k_{obs1}$, $k_{obs2}$, and $k_{obs3}$, corresponding to the fastest, intermediate and slowest exponential phases observed over 1000 s. These are atypical reaction transients not predicted for a cofactor independent enzyme with one active site, acting on a single substrate to yield a single product. For reactions of this nature one would typically expect to observe a single exponential curve consistent with first order kinetics.[357,358] Single turnover experiments can be used to simplify the interpretation of multiple-turnover data, as they enable direct interrogation of the rate of the chemical step of a reaction.[359] The appearance of three distinct phases within the AbyU single turnover experiment indicates the formation of three distinct species during a single catalytic cycle. In other transient kinetic analyses, multiphasic single turnover transients are usually indicative of a reaction proceeding *via* a stable transition state or multiple chemical reactions.[349,360,361] However, in this case, the detection of any viable transition state, according to the QM/MM data presented by Race and co-workers,[224] would not be feasible due to the femtosecond timescale observed during the asynchronous concerted DA reaction. Irrespective of this, the detection of intermediates in any of the hypothetically possible reaction pathways, whether concerted, biradicaloid or stepwise, would not be feasible due the timescales involved. Typical neutral [4 + 2] cycloaddition reactions can be observed both experimentally and theoretically to produce products in approximately 1 μs, whilst slower, ionic reactions are observed to produce products in under 1 ms,[362] with this latter rate still too quick for delineation by SF methods. Given that this reaction proceeds in a neutral system, that the diene is asymmetric, and asymmetric reactions typically proceed orders of magnitude faster[363] than the symmetric DA reactions described by Rivero *et al*.,[362] and the preceding QM/MM calculations, it can be safely determined that the three phases observed cannot be attributed to chemical transition states. Given that these distinct reaction species cannot, therefore, be attributed to transition state intermediates, two possible explanations remain. The first of these is that the enzyme substrate complex undergoes a large conformational change during the catalytic cycle. Such events are well documented in enzyme systems, for example in catalysis by uracil-DNA glycosylase amongst others.[364,365] In this instance however, this seems highly unlikely, as QM/MM simulations reveal very little conformational change in AbyU upon substrate binding,[224] or throughout the catalytic possess, and the robust β-barrel structures are known to possess inflexible protein folds.[351]

The third possible explanation for a multiphasic transient in a single turnover reaction between two distinct species, is the adoption of multiple substrate binding conformations within a single enzyme active site. This is a phenomenon that has been observed, albeit infrequently, through the inspection of substrate bound enzyme complex structures,[366] by comparison of chemical shift data by NMR methods,[367] or during mutagenesis of enzyme active sites.[359] This later case is elegantly illustrated by studies of Morphinone reductase, where a single active site mutation, removing a hydrogen bond to the substrate, facilitated the binding of NADH in multiple alternative conformations. This led to the detection of four separate binding conformations, equating to four discernible reaction rates during catalysis. Docking studies performed in the original AbyU paper show four potential docked poses of the substrate within the enzyme active site, of which one was postulated to be non-productive, one moderately productive and two highly productive, but with one pose possessing a higher energy barrier to reaction than the other. Given these data, one plausible explanation for the observed transients is that there are multiple binding conformations of **9** within the active site of AbyU. There is evidence in these transients of at least three different binding conformations, although there may be more which are not distinguishable, with rate constants needing to be roughly an order of magnitude in difference to be distinguished by SF.[359] Given the mechanism for enzyme catalysed DA reactions appears to be predicated on conformational restriction of the substrate (hypothesised for SpnF, MalC, AbyU, PyrI4), this lack of specificity is perhaps unsurprising and would not necessarily be expected to significantly hinder catalysis.

### 5.4.5 Single Turnover Studies of the AbyU Catalysed conversion of 9 to 10 With Respect to Substrate Concentration

In order to further investigate the catalytic mechanism of AbyU, multiple kinetic transients were obtained at a variety of concentrations of **9,** which was measured retrospectively using the starting absorbance recorded and the measured extinction coefficient for this substrate. Each syringe volume of reactants in the stopped-flow apparatus enabled the collection of 4 – 7 kinetic transients at each concentration of **9**. By fitting each transient to Equation 5, the rate of $k_{obs1}$ and $k_{obs2}$ with respect to [**9**], as well as the rate of the final linear phase, corresponding to a rough approximation of $k_{obs3}$, could be determined. Additionally, the amplitude of the faster exponential phase was determined ($A_1$), as was the amplitude of the slower exponential phase ($A_2$).

In order to analyse the change in $k_{obs1}$ with respect to [**9**], each set of repeats was plotted against [**9**] and the linear trend determined (Figure 5.17a, Equation 1). The same process was repeated for $k_{obs2}$ (Figure 5.17b). In this case the data were fit to a substrate inhibition curve (Equation 8). In addition, the rate of reaction observed in the initial linear phase of $k_{obs3}$ can be seen to increase in a linear manner with respect to [**9**] (Figure 5.18). This relationship can be fit to Equation 9 to obtain an approximation of the on and off rates observed in this portion of the reaction.

Equation 8:

$$\mathrm{Y} = \frac{V_{max} \times \mathrm{X}}{K_m + \mathrm{X} \times \left(1 + {}^{\mathrm{X}}/_{K_i}\right)}$$

Equation 9:

$$k_{obs} = k_{\mathrm{obs3\_}on}[\mathrm{S}] + k_{\mathrm{obs3\_}off}$$

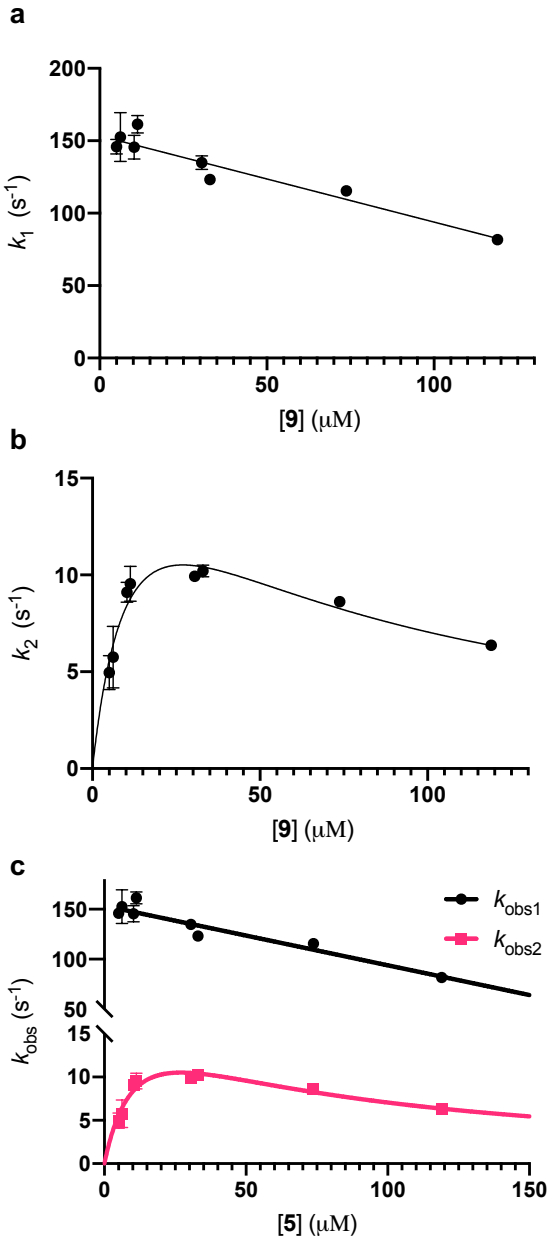



**Figure 5.17 Analysis of the rate constants observed during a single turnover of substrate 9 by AbyU.** Each data point is an average of 3-6 repeats and the error bars are standard deviation from the mean. a) The change in $k_{obs1}$ with respect to the concentration of 9. The data are fitted to a straight line (Equation 1). b) The change in $k_{obs2}$ with respect to the concentration of 9. The data are fitted to a substrate inhibition curve (Equation 8). c) $k_{obs1}$ and $k_{obs2}$ fitted on the same graph for direct comparison.

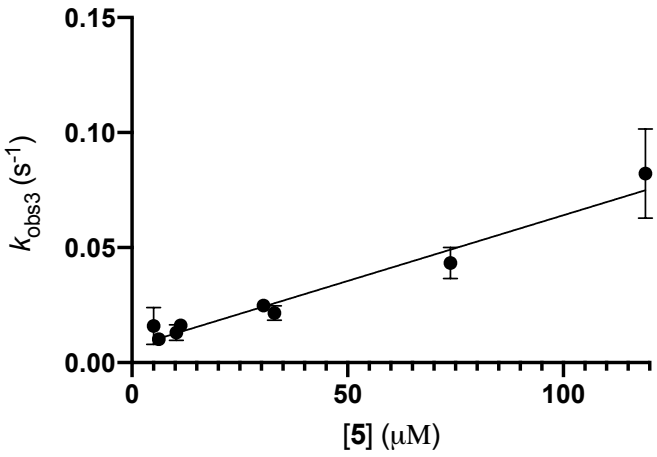



**Figure 5.18 Substrate Turnover in the Final Phase of a Single Turnover Reaction.** Each data point is an average of 3 – 6 data points and the error bars are standard errors from the mean. The initial rate of $k_{obs3}$ is fit to a straight line, the background rate of reaction removed, and plotted against the concentration of substrate **9**. The data are then fit to Equation 9, where $k_{obs3\_on} = 0.6 \pm 0.04$ $mM^{-1}$ $s^{-1}$ and $k_{obs3\_off} = 0.007 \pm 0.002$ $s^{-1}$.

The amplitude of each phase as a function of substrate concentration can also be investigated. This allows the determination of the binding constants associated with each phase, as well as the quantity of substrate needed to saturate a given binding conformation. The change in $A_1$ in relation to [**9**] was determined and the resulting data fit to a rectangular hyperbola (Equation 10, Figure 5.19a). The change in $A_2$ in relation to [**9**] was also determined. It was not possible to reach a concentration of substrate sufficient to saturate this binding conformation, but the data available fit best to a sigmoidal curve (Equation 11, Figure 5.19b). Intriguingly, if the amplitudes in the fast phase and the slow phase are combined and then plotted as a function of [**9**], we see that they fit to a linear regression that comprises ~68 % of the total substrate present in the reaction (Figure 5.19c). It appears as though the $A_2$, almost non-existent at very low concentrations of **9**, compensates for $A_1$ as it begins to saturate, and $A_2$ begins to predominate at concentrations of **9** above 58 μM.

Equation 10:

$$\mathrm{Y} = \frac{A_{lim} \times \mathrm{X}}{K_d + \mathrm{X}}$$

Equation 11:

$$\mathrm{Y} = \frac{A_{lim} \times X^h}{{k_{half}}^h + X^h}$$

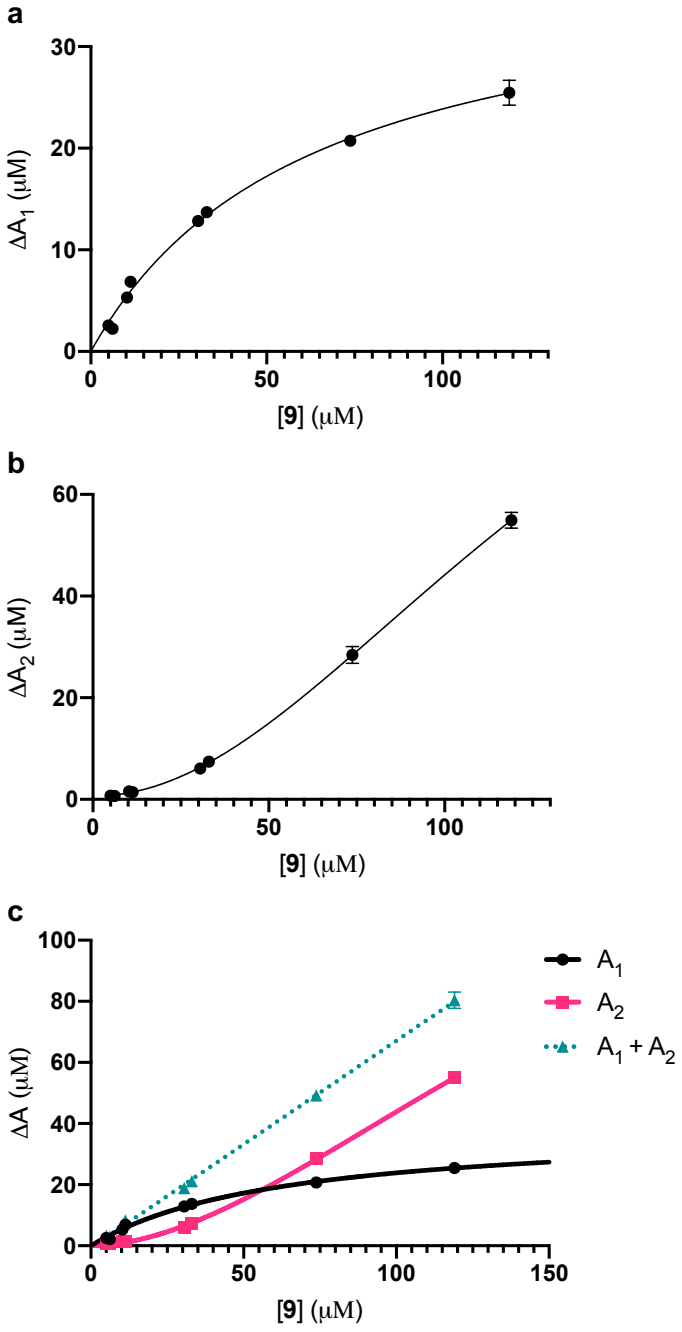



**Figure 5.19 Analysis of the amplitude change observed in each phase during a single turnover of substrate 9 by AbyU.** Each data point is an average of 3-6 repeats and the error bars are standard deviation from the mean. a) The change in $A_1$ with respect to the concentration of **9**. The data are fitted to a rectangular hyperbola (Equation 10). b) The change in $A_2$ with respect to the concentration of **9**. The data are fitted to a sigmoidal curve (Equation 11). c) $A_1$ and $A_2$, along with the amplitudes of $A_1$ and $A_2$ added together, and fit to Equation 1, on the same graph for direct comparison.

### 5.4.6 Interpretation of the Single Turnover Rates and Amplitudes as a Function of Substrate Concentration.

Analysis of $k_{obs1}$ and $k_{obs2}$ as a function of substrate concentration reveals a further layer of complexity. As the concentration of **9** increases, the rates of $k_{obs1}$ and $k_{obs2}$ unexpectedly decrease. This is a strong indication that within each observable phase there exists multiple different binding conformations. As the concentration of **9** increases, weaker, less productive, binding conformations are manifest, giving the appearance of a slowing of the reaction rate; this occurs because it is only possible to isolate rates that differ by at least an order of magnitude by stopped-flow. Consequently, it is not possible to extract intrinsic rate constants for each binding conformation. We are, however, able to ascertain the relative binding affinities observed for two different binding conformations contained within $k_{obs2}$. The $K_m$, 13.2 ± 3.2 μM, gives us a measure of the binding affinity for the quicker conformation, whilst the $K_i$, 54.9 ± 14.5 μM, represents the binding affinity for a less productive conformation which only becomes apparent at higher concentrations of **9**. These different binding conformations, which only begin to saturate at higher substrate concentrations, do not necessarily represent completely unique poses, as demonstrated by Byrne *et al*., but could also represent a need for small rearrangements to occur in the enzyme active site prior to the adoption of a catalytically competent state. The standard deviations observed in such analyses are routinely high, due to the inherent level of complexity within the system.

Analysis of $A_1$ and $A_2$ again reveals layers of complexity. Whilst the relationship between $A_1$ and the concentration of substrate fits well to a classical rectangular hyperbola (Equation 10) and produce an observable $K_d$ of 62.5 ± 3.9 μM and an $A_{lim}$ of 38.8 ± 1.2 μM; it is known from analysis of $k_{obs1}$ that these values represent an average for all binding conformations adopted in the single extractable phase. $A_2$ cannot be saturated under the experimental conditions and instead exhibits a sigmoidal relationship, although the data is insufficiently comprehensive to confidently verify this. What is clear, however, is that the sum of $A_1$ and $A_2$ is a consistent proportion of the substrate present in the reaction. The slope of the line reveals that 67.8 ± 0.6 % of the substrate is converted to product in the phases represented by $k_{obs1}$ and $k_{obs2}$. This leaves approximately one third of the substrate available to bind in the conformation(s) represented by $k_{obs3}$. The constant nature of this binding conformation reveals that it likely possesses a weak binding affinity for AbyU, and rather is dependent on the conformation that

substrate **9** adopts in solution. Given the previously discussed ability of individual molecules of **9** to self-associate, and the highly flexible nature of **9**, one possible explanation is that *inter*- or *intra*-molecular interactions preclude ~1/3 of the substrate from binding in a productive conformation within the enzyme active site. Analysing the portion of $k_{obs3}$ measurable over 10 s (Figure 5.18) enables determination of an over-estimated approximation of the rate of binding and dissociation for substrate in this conformation. This reveals an approximate $k_{obs3\_on}$ rate of 0.6 ± 0.04 $mM^{-1}$ $s^{-1}$ and $k_{obs3\_off}$ rate of 0.007 ± 0.002 $s^{-1}$. Although we are unable to extract reliable rate constants for $k_{obs1}$ and $k_{obs2}$, determining an approximate $k_{obs1\_off}$ rate of 150 $s^{-1}$ would conservatively indicate affinity of over 4 orders of magnitude higher in $k_{obs1}$ as compared to $k_{obs3}$. Mechanistically, two potential outcomes are likely when substrate is bound in this non-productive conformation. The first is the slow conversion of **9** to **10** within the AbyU active site before release. Alternatively, the substrate may be required to dissociate before re-binding in a productive conformation. This would mean that the substrate is capable of acting as competitive inhibitor of self, binding in a conformation that is essentially unproductive. Such behaviour is also observed in dihydrofolate reductase, for example.[367]

### 5.4.7 Burst Phase Kinetic Characterisation of the AbyU Catalysed Conversion of 9 to 10

To complement the single turnover experiments outlined above, and to provide additional insight into the reaction mechanism employed by AbyU, burst phase kinetic data were obtained using SF. These experiments provide information about the enzyme reaction mechanisms in the approach to the steady-state equilibrium, as well as the steady-state itself. The behaviour of an enzyme during multiple turnover reactions can deviate from that established using single turnover experiments. Data obtained using this approach can, however, be more challenging to deconvolute.[368] Crucially, multiple turnover reactions provide information about product release, which is often rate limiting.

SF transients for the AbyU catalysed conversion of **9** to **10** were obtained over 10 seconds at 325 nm and analysed on plotted on both logarithmic and linear axes (Figure 5.20). A burst phase, approximately stoichiometric to the concentration of AbyU used in the reaction, is clearly visible in these data, prior to establishment of a steady-state equilibrium. Fitting analysis was performed with a sloping baseline model, which fits to the linear steady-state phase

(Figure 5.21). A single exponential did not satisfactorily fit the burst phase data and biased the linear portion of the fit resulting in elliptical residuals. Two exponentials provide a good fit to the burst phase, and enable the linear portion of the curve to be fit without deviation. Three exponentials provide no significant benefit to the fitting statistics and so was deemed to be an overfitting of the data. This further underlines the complexity of the reaction, as multiple burst phases are rarely observed in experiments of this nature. It also serves to reiterate that the delineation of steady-state reaction kinetics in insolation can obscure complex and mechanistically critical information about catalytic processes.

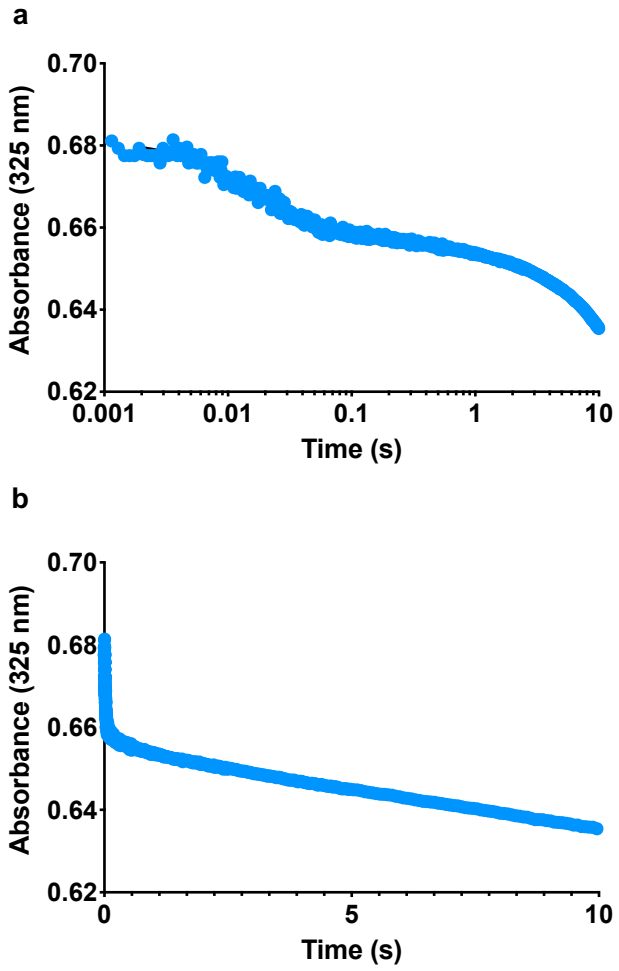



**Figure 5.20 Transients created by the reaction of 54 µM substrate 9 with 1 µM AbyU over 10 s.** The reaction was measured spectrophotometrically at 325 nm for 10 seconds. a) The transient displayed on a logarithmic scale. b) The same transient displayed on a linear scale.

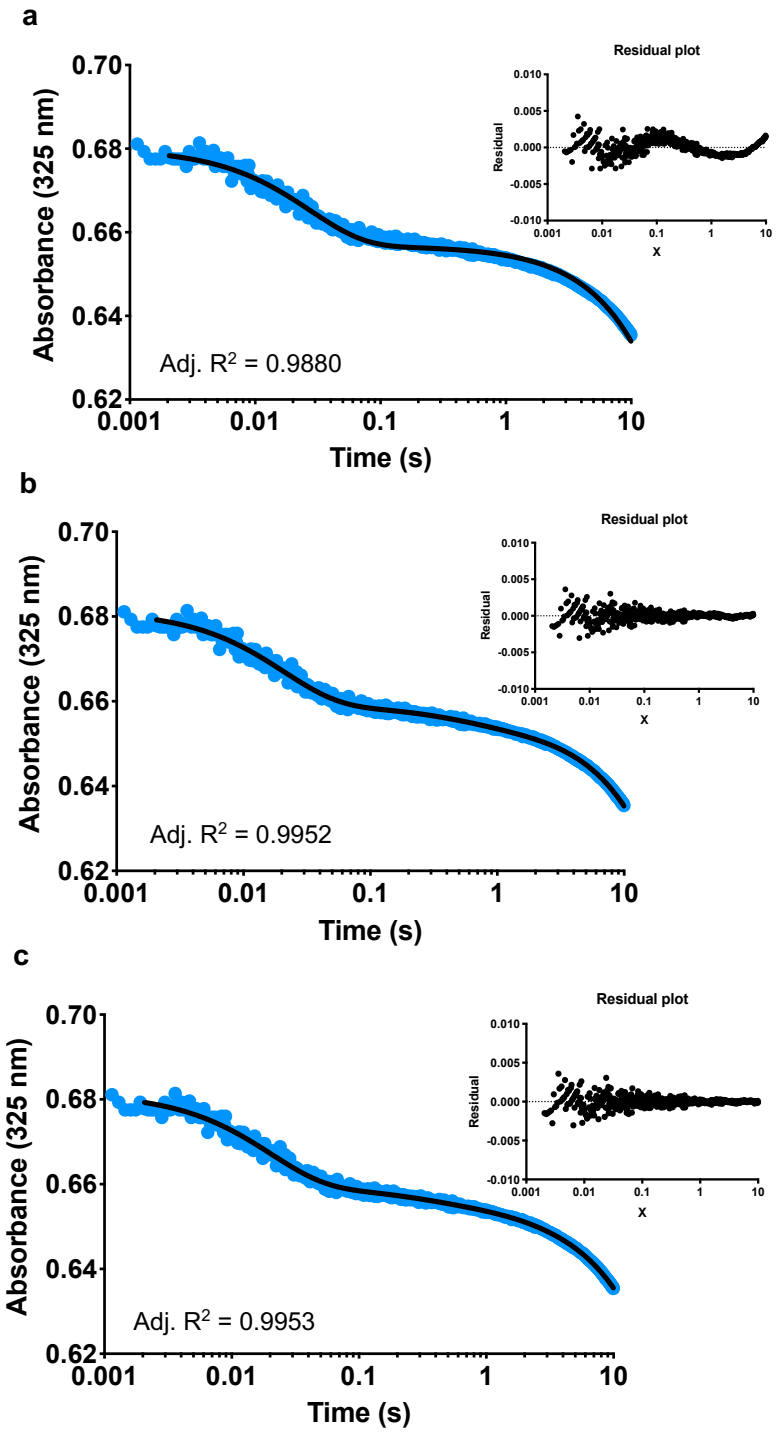



**Figure 5.21 Fitting analysis for the reaction of 54 µM Substrate 9 with 1 µM AbyU over 10 s.** The residual plots are inset and the adjusted $R^2$ values are given below the transients. a) The reaction transient fitted Equation 6. b) The reaction transient fitted Equation 5. c) The reaction transient fitted to Equation 7.

### 5.4.8 Analysis of Burst Phases with Respect to Substrate Concentration

As was the case with the single turnover experiments, the behaviour of the two burst phases and the linear steady-state phase were investigated as a function of substrate concentration. Burst phase kinetic transients were obtained at different concentrations of [**9**], repeated 3 – 6 times, and the rates observed for the fast exponential phase, $k_{b1}$, and slow exponential phase, $k_{b2}$, extracted. These results were then be plotted as a function of [**9**]. Due to the stoichiometric nature of these burst phases, the most likely explanation for origin is that product release is the rate limiting step, causing a population of [EP] complexes to rapidly accumulate before the steady-state is observed. For this reason, $k_{b1}$ has been fitted to a one-step binding equation (Figure 5.22a, Equation12), whilst the $k_{b2}$ data, which is representative of a two-step binding mechanism, was fit to Equation 13 (Figure 5.22b). The binding constants for each of these burst phases were then determined and are reported in Table 5.1.

Equation 12:

$$k_{obs} = k_{\mathrm{b1\_on}}[\mathrm{S}] + k_{\mathrm{b1\_off}}$$

Equation 13:

13.1 $$k_{obs} = {}^{1}/_{2}\,(p - q)$$

13.2 $$p = k_{\mathrm{b2\_on1}} \times [\mathrm{S}] + k_{\mathrm{b2\_off1}} + k_{\mathrm{b2\_on2}} + k_{\mathrm{b2\_on2}}$$

13.3 $$q = \sqrt{p^2 - 4r}$$

13.4 $$r = k_{\mathrm{b2\_on1}} \times [S] \times k_{\mathrm{b2\_on2}} + k_{\mathrm{b2\_off1}} \times k_{\mathrm{b2\_off2}} + k_{\mathrm{b2\_on1}} \times [S] \times k_{\mathrm{b2\_off2}}$$

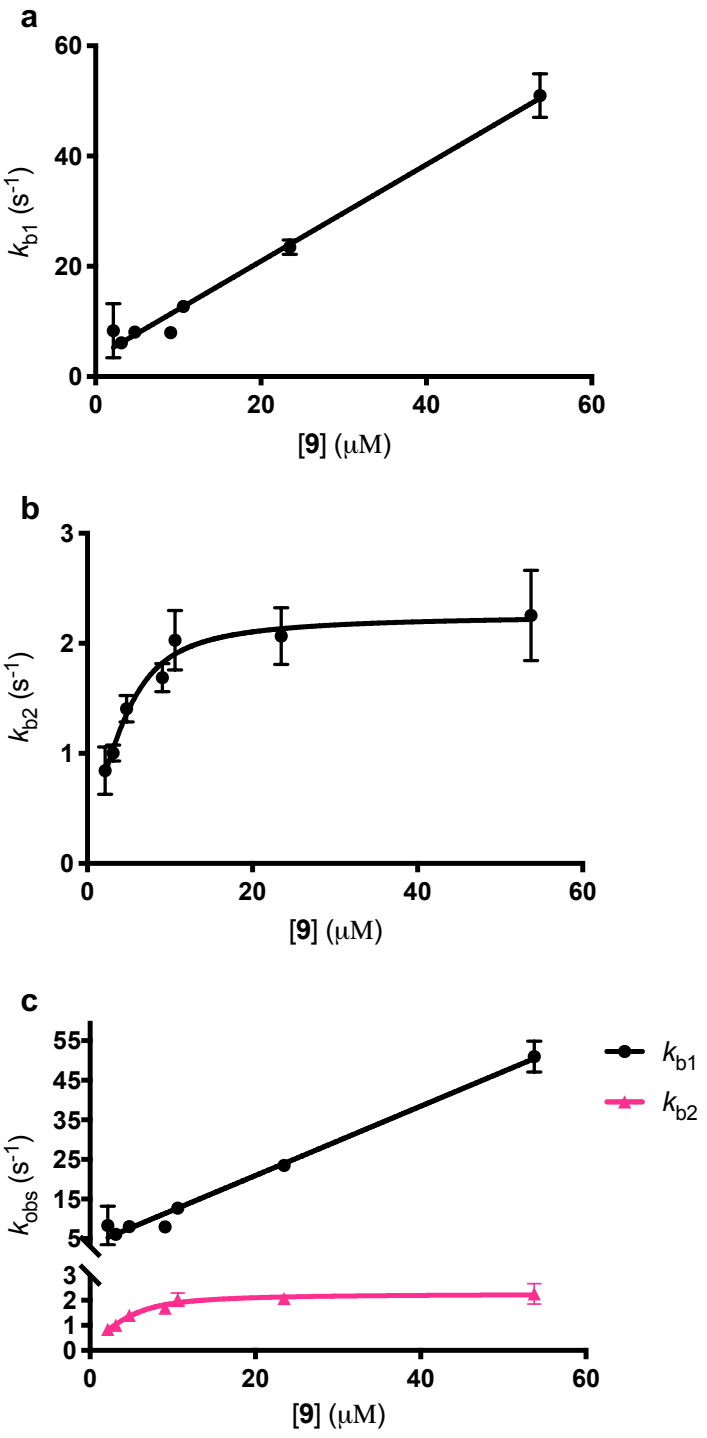



**Figure 5.22 Relationship Between the Rate Constants Observed and the Concentration of Substrate 9.** Each data point is an average of 3-6 repeats and the error bars are standard deviation from the mean. a) The change in $k_{b1}$ with respect to the concentration of **9**. The data are fitted to a one-step binding model (Equation 12). b) The change in $k_{b2}$ with respect to the concentration of **9**. The data

are fitted to a two-step binding model (Equation 13). c) $k_{b1}$ and $k_{b2}$ fitted on the same graph for direct comparison.

| | $k_{b1}$ | $k_{b2}$ |
|---|---|---|
| $k_{b\_on1}$ ($\mu M^{-1}$ $s^{-1}$) | 0.88 ± 0.03 | 0.36 ± 0.46 |
| $k_{b\_off1}$ ($s^{-1}$) | 3.4 ± 0.71 | 0.71 ± 0.62 |
| $k_{b\_on2}$ ($s^{-1}$) | - | 1.37 ± 2.38 |
| $k_{b\_off2}$ ($s^{-1}$) | - | 0.90 ± 2.29 |

**Table 5.1 Experimentally Determined Binding Constants for the Burst Phases Exhibited in the Turnover of 9 by AbyU.** $k_{b1}$ is fit to Equation 12 and $k_{b2}$ is fit to Equation 13.

### 5.4.9 Origin of Burst Phase Kinetic Transients

A small stoichiometric burst phase, as observed for the AbyU catalysed conversion of **9** to **10**, is often the result of the formation of a catalytic intermediate during an enzyme catalysed reaction. This means an intermediate must saturate before the steady-state can predominate. For the AbyU catalysed reaction, as it is possible to discern two distinct burst phases, there must be two distinct catalytic intermediates formed. As discussed previously, a large conformational change in AbyU during the catalytic step, or the formation of a chemical intermediate, are both implausible. The mechanistic implication of these data is thus that product release is rate limiting. This is perhaps unsurprising as the conformationally constricted product must by necessity dissociate from an enclosed, hydrophobic active site. This process is further regulated by lid opening, or the movement of the capping loop sequestering the product (highlighted in Figure 5.8). This loop region has been shown to be important in catalysis in this class of enzyme,[158] and so it is not surprising that it would play a central role in the catalytic cycle.

The presence of two distinguishable burst phases gives further credence to the hypothesis that multiple binding conformations are observed during catalysis. The faster of the two burst phases observed, $k_{b1}$, likely represents substrate binding in the amalgamation of subtly different conformations detected in $k_{obs1}$. The data fit well to a one-step binding model, allowing us to extract on and off rates for this conformation in a competitive environment, as reported in Table 5.1. The rates observed in $k_{b2}$, however, fit to a two-step binding model. This likely represents

substrate binding in the conformations represented by $k_{obs2}$, and the two-step behaviour could be a consequence of the substrate undergoing a modest conformational rearrangement within the active site before adoption of a catalytically competent pose is adopted. The burst phase kinetic transients obtained (Figure 5.20) also enable the determination of the Michaelis constants for the AbyU catalysed reaction (Figure 5.23; Table 2). These data provide information on substrate specificity and the rate of catalysis under equilibrium conditions. Steady-state values obtained can be complex functions comprised of multiple contributing steps.[359]

Equation 14:

$$Y = \frac{V_{max} \times X}{K_m + X}$$

These analyses yield a $k_{cat}$ of 6.94 ± 0.30 $min^{-1}$ and *K*m of 9.9 ± 1.1 μM. These values deviate from those previously reported by Byrne *et al*.[224] Although this discrepancy is not easily resolved, the experiments performed previously were conducted using a questionable protocol. The $k_{cat}$ determined here is ~230 times slower than the initial rate of $k_{b1}$, giving a measure of the limit in rate that product release imposes.

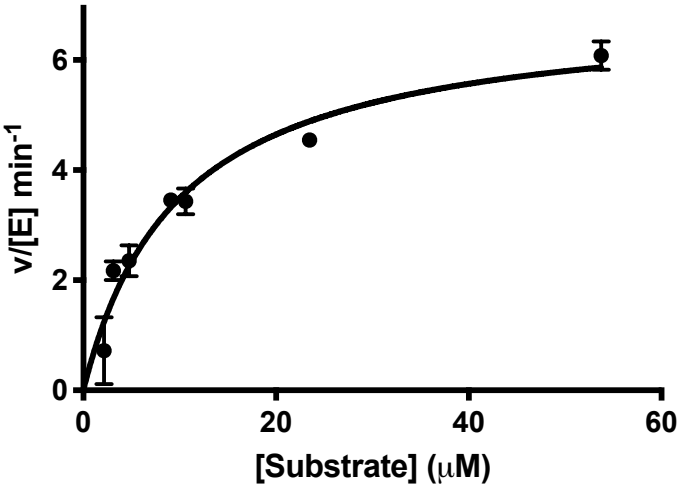



**Figure 5.23 The steady-state reaction between AbyU and substrate 9.** The data are fit to a Michaelis-Menten curve (Equation 14). Each data point is the average of 3-6 repeats and the error bars are standard deviation from the mean.

| Michaelis constant | Value |
|---|---|
| $k_{cat}$ ($min^{-1}$) | 6.94 ± 0.30 |
| $K_m$ (μM) | 9.9 ± 1.1 |
| $k_{cat}$ / $K_m$ ($\mu M^{-1}min^{-1}$) | 0.70 ± 0.11 |

**Table 5.2 Steady-state kinetic parameters for the reaction of AbyU and substrate 9.**

### 5.4.10 Preliminary Investigations into the Effect of Temperature on the AbyU-Catalysed Cycloaddition reaction

Further investigation into the catalytic mechanism employed by AbyU could entail performing an analysis of the single turnover reaction at different temperatures. Preliminary analysis at 5 °C reveals a fourth extractable exponential phase present in the transient (Figure 5.24), corroborating the prior hypothesis that the extractable rates actually contain a multiplicity of catalytic conformations. This additional extractable phase appears to come out of $k_{obs2}$, and likely corresponds to the second phase with a binding affinity similar to that of the measured $K_i$ for $k_{obs2}$ at 22 °C. Intriguingly, we are also able to observe a significant increase in the proportion of substrate turned over by AbyU in 1000 s compared to what is seen at 22 °C (Figure 5.24c and Figure 5.25). Although the rates observed for $k_{obs1}$, $k_{obs2}$ and $k_{obs3}$ appear to decrease, as would be expected, the amplitudes observed in each phase increase. This corresponds to an additional 33 % turnover over the full timescale and enables all the substrate present in the reaction to be cyclised. Mechanistically, this implies that the substrate has more chance to bind correctly and is less likely to bind in an inhibitory manner. This raises the interesting possibility that AbyU could be a more effective catalyst in cold environments, something that is not entirely surprising for an enzyme isolated from the deep sea.

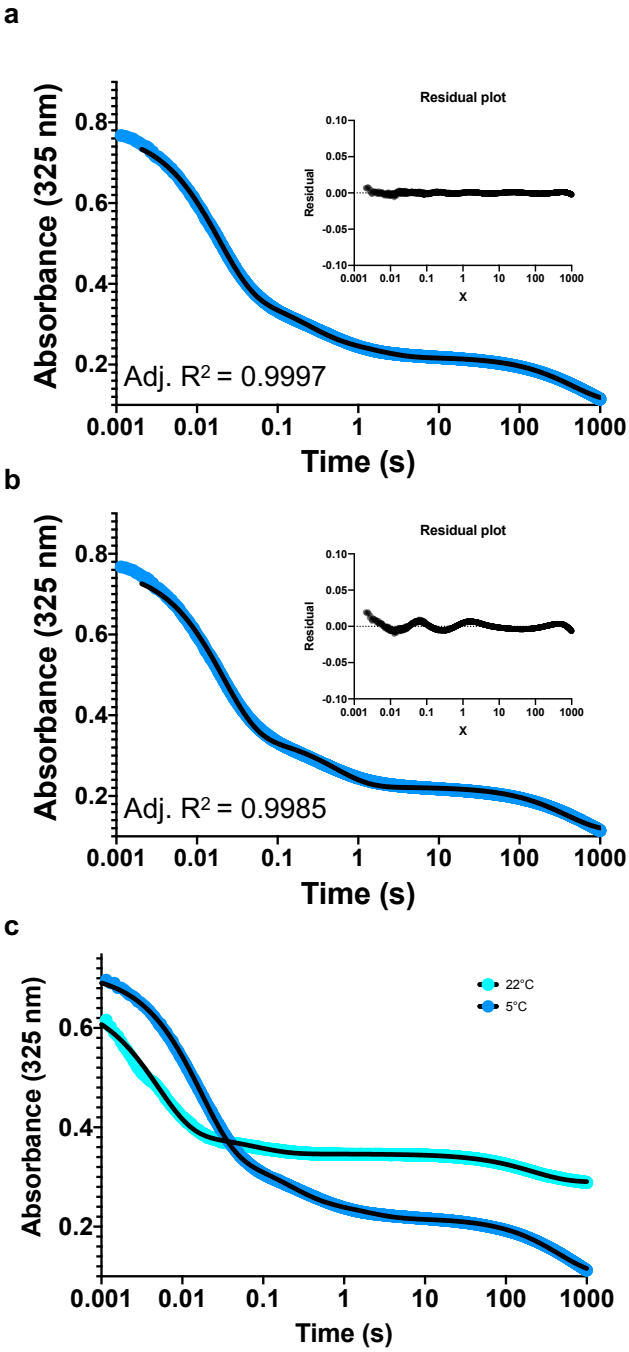



**Figure 5.24 Transients and fitting statistics for a single turnover of 9 by AbyU at 5°C.** a) The reaction transient fitted to Equation 4. b) The same reaction transient fitted to the same equation with three exponential terms used to fit the transients obtained at 22 ºC (Equation 3). In comparison to experiments performed at room temperature an additional phase is observed, becoming apparent around $k_{obs2}$. c) A comparison of transients obtained at 22 ºC and 5 ºC over 1000 s.

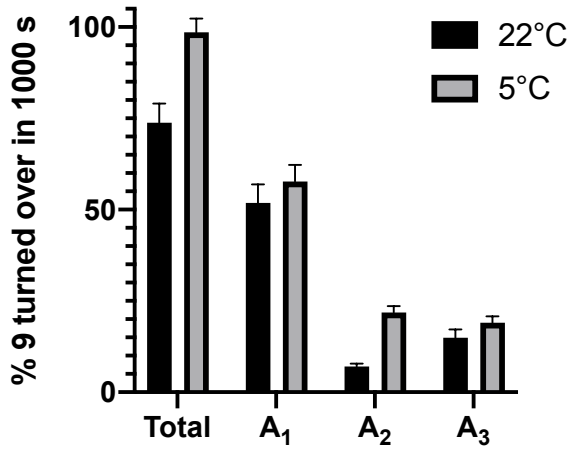



**Figure 5.25 The proportion of 9 turned over in 1000 s at 22°C and 5°C.** A comparison between the turnover observed in each extractable exponential phase observed at both 22°C and 5°C, and in the total turnover over 1000 s. $A_2$ at 5 °C is a composite of the amplitude observed in both phases extractable between $A_1$ and $A_3$. Each bar is an average of 6 results at 5°C and 3 results at 22°C.

## 5.5 Discussion

### 5.5.1 Analysis of AbyU's Synthesised Substrate Analogue 9

During this study, substrate **9** was found to present a number of challenges when used in *in vitro* enzyme assays. In contrast to other putative DA substrates, **9** is a tractable synthetic target. However, the potential for polymerisation of this compound during its preparation necessitates the introduction of a tetronate ring methyl protecting group to stabilise this compound in solution. **9** was also found to be challenging to work with at lower concentrations due to an inability to produce stable and appropriately concentrated stock solutions, and the propensity of **9** to adhere to plastic labware. These issues translate into significant errors during handling. Fortunately, however, during the experiments outlined herein, appropriate handling protocols were established to circumvent these effects. DLS analysis of **9** does demonstrate that this compound can self-assemble into micellular like structures at concentrations > 500 μM, however, this is above the threshold required to conduct the experiments described in this thesis.

### 5.5.2 Mechanistic Insights into AbyU Catalysis

Together, the single turnover and burst phase kinetic data present herein represent the first detailed mechanistic investigation into the AbyU catalysed cyclisation of **9**, and indeed the first such analysis of a [4 + 2] cyclase. These data reveal a minimum of three distinct binding conformations of substrate within the enzyme, as established from the single turnover data. Analysis of the rates of $k_{obs1}$ and $k_{obs2}$ with respect to substrate concentration reveal an additional layer of complexity. Decreasing rates of reaction imply that these phases comprise multiple binding conformations, with slower reacting conformations becoming more prevalent at higher concentrations of **9**. In order to further investigate the apparent multiplicity of binding conformations, computational simulations were performed by Mr. Samuel Johns at the University of Bristol (Figure 5.26). The substrate was docked into the active site in a variety of conformations and molecular dynamics (MD) simulations were performed to ascertain the most representative pose for each docked structure. This revealed four possible general conformations of the substrate within the active site (A-D), which are consistent with the findings previously presented by Byrne *et al*. MD was then used to determine binding energies for each pose, and QM/MM simulations were performed to calculate the reaction free energy. These analyses reveal three

distinct binding poses with relatively similar binding energies (A, B and D). Whilst substrate bound in conformations A and B have relatively similar activation free energies, the reaction free energy required in conformation D is significantly larger. An additional small free energy penalty is expected for substrate bound in conformation B because of the longer distances between the diene and dienophile observed in MD simulations. This corresponds well to the three distinguishable binding conformations represented by $k_{obs1}$, $k_{obs2}$ and $k_{obs3}$, with substrate bound in conformation D experiencing a significant barrier to catalysis that is possibly too great to overcome.

The burst phase kinetic data provides additional insight into the catalytic mechanism of AbyU and corroborates the single turnover and simulation data. These data provide evidence for multiple catalytic binding conformations in the burst phases, with two distinct reaction intermediates forming prior to the steady-state. The accumulation of these intermediates indicates that product release is rate limiting, a conclusion that makes mechanistic sense when the structure of the enzyme is considered. The fit of $k_{b2}$ to a two-step binding model is most revealing aspect of the burst phases. This implies that substrate binding in this conformation is undergoing a significant rearrangement in the active site before catalysis can proceed. Whilst the MD simulations performed have precluded the possibility of the substrate flipping about a horizontal axis, i.e. a transition from C or D to A or B, only a minor dissociation would be required for the substrate to rotate about a vertical axis. With this in mind, the most likely hypothesis for the existence of a two-step binding model is that substrate binding in conformation B is, in fact, undergoing a slight dissociation and re-binding in conformation A prior to catalysis.

One intriguing question is as to why there is no manifestation of substrate bound in conformation D, that is those conformations represented by $k_{obs3}$. The rate of reaction in $k_{obs3}$ is so slow that cyclisation in this conformation would not necessarily be evident over the timeframes used in this study. However, if we were to assume that ~1/3 of the substrate would bind in conformation D in each reaction cycle, as shown in the single turnover experiments, the AbyU active sites would rapidly become inhibited and the steady-state rate would be an approximation of $k_{obs3}$. Although they do not necessarily agree with the binding energies determined for each pose, the calculated approximations of the $k_{obs3\_on}$ and $k_{obs3\_off}$ rates of $0.6 \pm 0.04$ $mM^{-1}$ $s^{-1}$ and $0.007 \pm 0.002$ $s^{-1}$, respectively, provide further insight. Despite being determined in a non-competitive single turnover environment, they are around three orders of magnitude smaller than either of the on or off rates determined in each burst phase. In a competitive, multi-

turnover, scenario, binding in the conformations represented by $k_{obs3}$ would be outcompeted by substrate binding in conformations A and B. The modelling data only take into account the affinity for the pose once it is inside the active site.

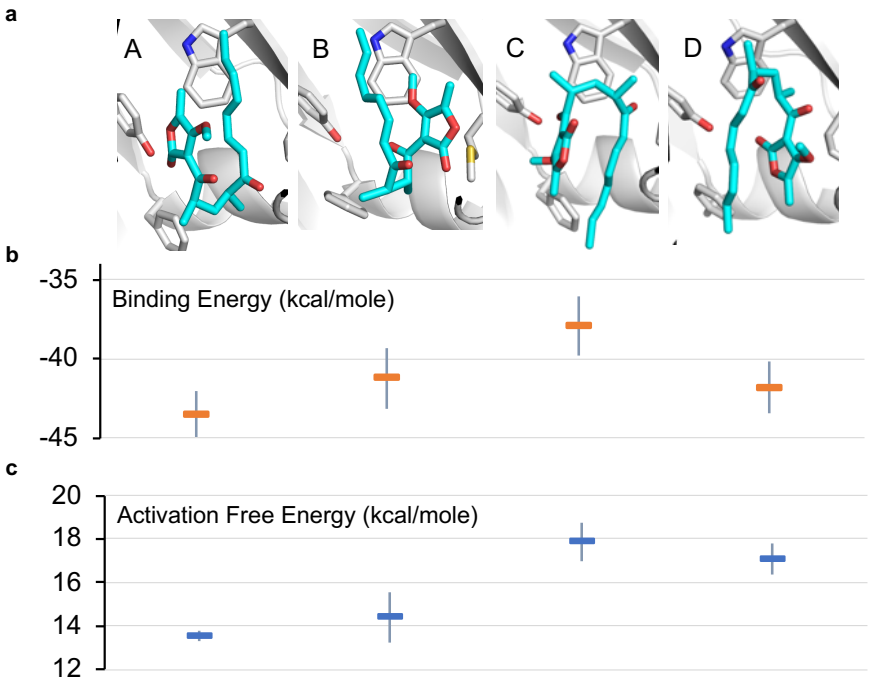
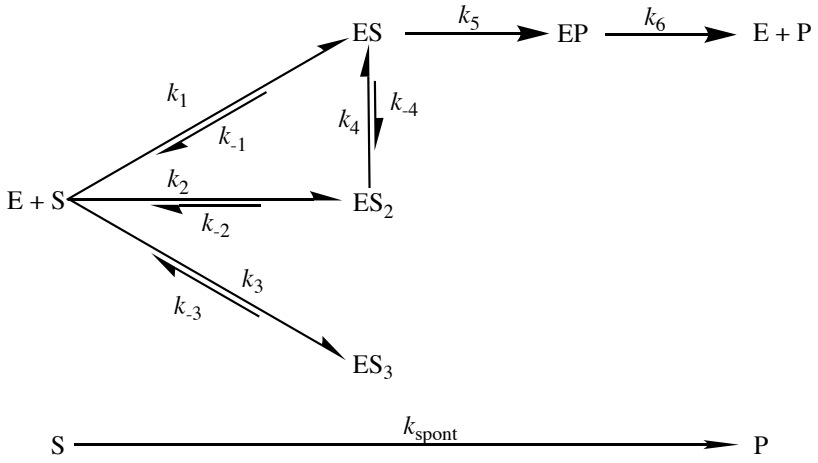



**Figure 5.26 MD and QM/MM Classification of Different Substrate Binding Conformations.** a) The representative structures from molecular dynamics (MD) simulations of substrate **2** in each of the 4 binding modes obtained from docking. b) Average binding energy for each mode (as calculated using the MM-GBSA approach using 1000 snapshots of each mode taken from 10 independent MD runs). c) Predicted Diels-Alder activation free energies for each pose (obtained from 10 independent QM/MM umbrella sampling molecular dynamics simulations using DFTB/ff14SB, with a 7.32 kcal/mole correction for each to M06-2X/6-31+(d,p) level). Standard deviations are indicated for binding and activation free energies. Figure and Figure Legend courtesy of Mr. Samuel Johns, University of Bristol.

The kinetic transients obtained at 5 °C appear to support the theory that, at 22 °C, substrate is rapidly bound in non-viable conformations. Although the initial rate of turnover at 5 °C is slower than that observed at 22 °C, the amplitudes are increased, particularly with respect to $k_{obs2}$. This appears to show that by slowing the rate of binding, AbyU is able to more effectively rearrange the substrate into a productive conformation before the substrate is fully bound within the active site. It is likely that a complex system of small interactions between substrate

**9** and the loop region and / or the initial residues in the cavity of the barrel act to ensure correct turnover of the substrate under physiological conditions. For an enzyme isolated from a bacterium found at the bottom of the ocean, this could well be much closer to 5 °C than 22 °C. These initial findings need to be verified through further experiments before concrete conclusions can be drawn.

The combination of the transient approaches applied during this study, coupled with *in silico* analysis, enables the construction of a reaction scheme for AbyU catalysis (Scheme 1). Within this scheme the three different observable binding conformations are represented by $ES_1$, $ES_2$ and $ES_3$. These correspond to the binding conformations identified in Figure 5.24 as A, B and D, respectively. Molecules of substrate forming an ES complex are able to undergo rapid catalysis, yielding an enzyme-product complex, EP. The product **10** is then held within the active site, enclosed within the hydrophobic pocket by the capping loop, before being slowly released. Substrate binding to form $ES_2$, likely in conformation B, is in fact rearranged to form ES, before cyclisation occurs. Substrate binding to form $ES_3$ experiences a large barrier to catalysis meaning that most substrate is likely to slowly dissociated prior re-binding in a productive conformation. Simultaneously, free substrate will undergo spontaneous cyclisation to form product at rate $k_{spont}$. The combination of burst phase kinetic data and an overestimated approximation for the on and off rates for substrate binding in a non-productive mode, as illustrated by binding mode D, from the single turnover experiments allow us to determine binding constants for $ES_1$, $ES_2$ and $ES_3$ (Table 5.3). It is important, however, to recognise that this scheme is likely to represent a simplification of the actual reaction mode. The single turnover experiments reveal that each of ES and $ES_2$ represent an average of multiple, subtly different, substrate binding conformations, and hence the on and off rates depicted are a composite of a multitude of different rates with closely related affinities. The steady-state rate provides an approximation for product release ($k_6$), whilst cyclisation occurs extremely rapidly once the substrate is bound ($k_5$).




**Scheme 5.1 Proposed reaction scheme for the AbyU catalysed intramolecular DA reaction.** There are three sets of binding conformations that can be distinguished experimentally, denoted ES, $ES_2$ and $ES_3$. The intrinsic rate constants calculated are given in Table 5.3.

| Parameter | Rate Constant |
|---|---|
| $k_1$ ($\mu M^{-1}$ $s^{-1}$) | 0.88 ± 0.03 |
| $k_{-1}$ ($s^{-1}$) | 3.41 ± 0.71 |
| $k_2$ ($\mu M^{-1}$ $s^{-1}$) | 0.36 ± 0.46 |
| $k_{-2}$ ($s^{-1}$) | 0.71 ± 0.62 |
| $k_3$ ($\mu M^{-1}$ $s^{-1}$) | ≈ 5.7 x$10^{-4}$ ± 3.8 x$10^{-5}$ |
| $k_{-3}$ ($s^{-1}$) | ≈ 6.9 x$10^{-3}$ ± 1.9 x$10^{-3}$ |
| $k_4$ ($s^{-1}$) | 1.37 ± 2.38 |
| $k_{-4}$ ($s^{-1}$) | 0.90 ± 2.29 |
| $k_5$ ($s^{-1}$) | N.D. (very quick) |
| $k_6$ ($s^{-1}$) | ≈ 0.12 ± 0.05 |
| $k_{spont}$ ($\mu M^{-1}$ $s^{-1}$) | 1.25 x$10^{-3}$ ± 1.7 x$10^{-5}$ |
| $K_1$ (μM) | 3.86 ± 0.94 |
| $K_2$ (μM) | 1.97 ± 4.24 |
| $K_3$ (μM) | ≈ 1.2 x$10^{-5}$ ± 4.1 x$10^{-6}$ |
| $K_4$ | 0.66 ± 2.81 |
| $K_{app}$ (μM) | 1.3 ± 8.4 |

**Table 5.3 Experimentally determined binding constants for the cyclisation of 9 by AbyU.** $K_1 = k_{-1} / k_1$ and $K_2 = k_{-2} / k_2$. $K_3$ and $K_4$ are calculated in the same way. $K_{app} = K_2 * K_4$.

### 5.5.3 Consequences for the Exploitation of DA Biocatalysis

The unexpected level of complexity uncovered by the studies of AbyU reported herein has significant consequences for the development of an industrially relevant DAases. Attempts at engineering this class of enzyme should take into account the multiple binding conformations observed, as mutations within the barrel or the loop region could have the unintended consequence of impacting substrate binding affinity, particularly in orientations C and D. Increasing the specificity for optimally bound substrate could, however, be one opportunity for engineering attempts. The highly significant rate-limiting step in the catalytic cycle, the opening of the lid-like loop region allowing the release of product **10**, is a barrier to optimal catalytic proficiency and presents another opportunity for rational engineering. Whilst removal of the loop region in its entirety is not likely to provide benefit,[158] mutagenesis could increase the *in vitro* catalytic rate. The work further describes possible reaction limits imposed by substrates in this type of system. Enzymatic catalysis cannot optimally occur at high substrate concentrations, with the large higher order structures observed by DLS likely to seriously inhibit anything apart from small scale, low concentration catalysis.

### 5.5.3 Summary

This work constitutes the first investigation of its kind into the mechanism of a [4 + 2] cyclase, and similarly the first detailed kinetic analysis of a β-barrel enzyme of any description. The results show that this class of enzyme, presumed to function *via* a simple catalytic mechanism of conformational restriction, is, in fact, a deceptively complex biocatalytic system. This work shows that the current practice of characterising putative DAases using steady-state methods is likely to conceal a number of catalytic nuances which must be understood before efforts to rationally redesign these enzymes for industrial applications can succeed.

# Chapter 6

# Mechanistic and Biophysical Studies of AbyU at High Pressure

## 6.1 Introduction

### 6.1.1 Determining the Effect of Pressure on Proteins

The effect of pressure on biomolecular systems has been studied since P. W. Bridgeman published a short article about the effect of high pressure (HP) regimes upon Albumin, specifically that pressure drives coagulation.[369] The effect of pressure on enzyme structure and function is still not fully understood, despite pressure being a ubiquitous thermodynamic property. The effect of pressure on the stability and activity of several model enzymes was studied from the 1960s to 1990s.[370–374] In general, these studies sought to determine the basic influences of pressure on proteins. The development of HP UV-vis spectroscopy facilitated many of these investigations, allowing the *in-situ* determination of pressure induced protein denaturation. For example, Zipp and Kauzmann were able to determine the conditions for the denaturation of sperm whale metmyoglobin in relation to a combination of temperature and pressure.[375]

Many of the early studies into the effects of pressure upon proteins were hindered by a requirement for access to expensive specialist equipment and were mostly of limited scope. Nevertheless, several significant advances were made. In the 1970s it was determined that high pressure regimes could cause multimeric enzyme complexes to dissociate, thereby inhibiting their function,[373] whilst the effect of pressure upon the folding landscape of enzymes was first discussed by Walter Kauzmann in Nature in 1987.[372] Similarly, the effect of pressure on more complex biological systems was also investigated, for example, Robb and Clark's study of the growth of hyperthermophiles at HP. This study demonstrated significant benefits to culturing these microorganisms at HP environments, postulated to be due to enzyme adaptation to pressures in the deep sea.[370] Together, these investigations established that pressure has the potential to significantly influence the structure and behaviour of biological molecules and

macromolecular systems. Pressures relevant to cells, proteins and biochemical reactions are generally thought to be in the range of 1 bar – 10 kbar, whilst it is accepted that pressures up to 20 kbar do not impact the primary structure of biological molecules such as proteins, peptides and lipids.[376] For context, the deepest point in the ocean, the Mariana trench, which represents the point of highest pressure on earth, has a maximum depth of 11,000 m, equating to a pressure of 1,100 bar.

More recently, detailed molecular level studies of the influence of pressure on enzyme structure and mechanism have been largely restricted a small number of research groups internationally. These studies have generally used specialised biophysical techniques, including high pressure nuclear magnetic resonance (HP-NMR) spectroscopy, pressure jump studies and, most recently, high pressure stop-flow (HP-SF).[377] Kremer *et al.* have been able to use pressure to stabilise excited protein folding states and then interrogate them using HP-NMR.[378] Hay *et al*. were able to interrogate electron transfer in redox proteins *via* HP-SF, sampling electron transfer rates in different protein folding states stabilised by HP regimes.[379] High pressure regimes have, of course, been far more extensively studied within synthetic chemistry, and have frequently been shown to have a catalytic effect, driving many synthetic chemical processes, as well as promoting isomerisation and selective polymerisation.[380,381]

### 6.1.2 The Effect of Pressure on Enzymatic Catalysis

HP enzyme catalysis studies have largely focused on well characterised model biocatalysts, including α-chymotrypsin,[382–384] and human cytochrome P450 by Hay *et al*.[379] These studies seek to investigate and explain the fundamental rules determining what effect pressure elicits upon an enzyme's catalytic performance and structural integrity.[383,385,386] These studies have shown that, generally, HP serves to strengthen hydrogen bonds within biomolecules, whilst destabilising hydrophobic reactions.[387] The secondary structure composition of an enzyme, and the structure-function relationships necessary for catalytic competency, have been shown to play a major role in defining the capability of enzymes to recover after pressurisation. For example, Araujo *et al*. showed that whilst Hsp70 is not impeded as an ATPase, the application of pressure to 3000 bar completely inhibits its ability to refold heat-denatured luciferase.[386] Cavities within the structure required to support substrate re-folding are proposed to be

vulnerable to pressure-induced denaturation, preventing the restoration of function following pressurisation.

### 6.1.3 The Effect of Pressure on Enzyme-Catalysed Chemical Reactions

HP regimes have been shown to decrease available reaction volume. This means that, in a synthetic context, HP drives reactions in the direction of bond formation, as this decreases the volume required by the individual reactants, thus decreasing the overall volume of the system.[376,380,381] More broadly, if a reaction transition state has a smaller volume than that of the reactants, HP tends to promote the reaction rate. This provides an explanation as to why high pressure regimes are often used to drive Diels-Alder reactions, with a net reduction in product volume vs. that of the reactants. This is consistent for both *inter*- and *intra*-molecular processes.[388] The effect of pressure on enzyme-catalysed chemical reactions is decidedly more complex, as illustrated by the work of Butz *et al*., who investigated the reversible conversion of fumarate to L-malate by fumarase.[389] When fumarase binds these substrates, this results in a decrease in the volume of the reaction transition state, but an increase in the volume of the enzyme-fumarate or enzyme-malate complex. Binding further displaces water molecules bound to catalytic side chains, thus increasing the reaction volume. Thus, a volume minimum is observed during binding of either fumarate or malate, and a maximum observed as the chemical intermediate forms, as water is displaced, but prior to bond formation. Similarly, the proteolysis of ubiquitin by α-chymotrypsin has been investigated at HP. Even though proteolysis results in the breaking of bonds and a subsequent increase in reaction volume, the rate of α-chymotrypsin-catalysed proteolysis is greatly enhanced by a HP regime.[384] As ubiquitin may adopt a proteolysis-resistant folded form, or a highly susceptible unfolded form,[390] it is postulated to be readily unfolded by the application of pressure, as disordered proteins have a much smaller volume than ordered proteins. This enables a highly increased rate of reaction, despite subsequent volume increases as a result of proteolysis.

Broadly speaking there are three reasons an enzyme's activity may be enhanced by a HP regime. HP may drive a conformational change within the enzyme, or within the target protein itself (such as ubiquitin) that drives catalysis. HP may also affect the reaction mechanism by driving conformational changes that decrease the volume of the enzyme. This can result in both increased and decreased rates of reaction, particularly if it affects the rate limiting step. Thirdly,

HP can initiate a change in either the solvent, such as the pH or viscosity, or in the substrate, by modulating solubility or *inter*-molecule associations, thereby affecting either the enzyme or the chemical reaction.[391] When the effect of temperature on enzymatic catalysis is also taken into account, another set of considerations emerge. For example, HP regimes used to be thought to only have enhancing effects on monomeric enzymes, as the application of HP drives multimer disassembly in the majority of cases. This decreases the volume of the system by exposing hydrophobic surfaces.[376] It has since been shown, however, that some multimeric enzymes are stabilised or enhanced at high pressure, particularly at temperatures high enough to otherwise dissociate the protomer complex.[391,392] The ability of HP regimes to stabilise enzymes at HP comes about due to heat denaturation, and subsequent aggregation, which acts to increase the relative volume occupied by the biomolecule. This is a direct consequence of the formation of multiple, non-specific, interactions.[376] Enzyme catalysis, which is often accelerated by temperature increases, can therefore be accelerated further by a combination of high temperature and HP.[377]

More detailed mechanistic studies have made use of the HP-SF system developed by TgK Scientific. These studies have been used to both to investigate the model enzymes previously highlighted, and to explore enzyme catalysed hydrogen tunnelling.[393,394] Studies of model systems have yielded important insights; they are, however, limited in scope and generality, focussing predominantly on hydrolases and oxidoreductases.[391]

### 6.1.4 Investigating Protein Folding Using High Pressure

Although not the focus of this study, it is worth discussing, briefly, some of the studies that have successfully used HP-NMR to investigate protein folding and to identify intermediate folding states in proteins.[395,396] This technique enables the chemical shifts observed in individual main chain amide groups to be studied, thus providing information about any rearrangements occurring due to the HP regime. A sigmoidal, and therefore cooperative, relationship between individual groups is often observed and correlates to a conformational transition, or the movement of the folded state to a new, preferential minima within the protein energy landscape. This approach can be used to bias the formation of little seen yet physiologically relevant folding intermediates. For example, HP-NMR has been used to study the folding of prion proteins that are capable of propagating in a species-specific manner.[397] These experiments do, however, have a variety of drawbacks. Principally, this technique is only useful for the study

of small, globular and previously well characterised enzymes, such as ubiquitin.[398] It can also be extremely challenging to identify how macromolecular changes impact the secondary and tertiary structure of proteins, particularly if there has been no significant enhancement of an alternative conformer.

High pressure Fourier transform infra-red spectroscopy (HP-FTIR), provides far greater insight to the changes in secondary structure composition instigated by HP regimes. It has, however, only been sparingly used to study, for example DNA, or the formation of amyloid fibrils. This is due to the practical complexities of this technique.[385,399] It requires highly concentrated protein samples, or, ideally, a film such as is found in amyloidosis. $D_2O$ is mandatory as a solvent at HP regimes, and diamond anvil windows are sub-optimal for FTIR spectroscopy due to similar spectral characteristics to amide regions.[400] High pressure synchrotron radiation circular dichroism (HP-SRCD) is also capable of determining secondary structural changes *in situ* at HP. This technique possesses its own drawbacks, including the requisition of a synchrotron. Solvation in more benign phosphate-based buffers is possible, however, as is the use of less concentrated samples than in HP-FTIR, where samples concentrated up to 10 % are sometimes required.[399] Studies using this technique have, to date, been extremely limited, although it has been used to monitor excited folded states of myoglobin.[401]

### 6.1.5 Alternative Applications of High Pressure Enzymology

HP regimes and their effect on enzyme function have also been investigated by food scientists, and their findings are worth briefly outlining here. These studies have sought to determine catalytic function following high pressure processing (HPP), a technology emerging in the food industry as an alternative to commercial thermal processing techniques which are widely adopted. These studies have sought to apply a HP regime to an enzyme and then to observe structure and function when the enzyme has returned to ambient pressures, with *in situ* characterisation of little relevance.[402–404] These studies have confirmed that whilst some enzymes can be denatured by HPP, others can be moderately enhanced by it., including some antioxidant enzymes relied upon by the food industry to preserve some food stuffs. The lack of either *in situ*, or comprehensive, biochemical analysis, means that these studies are of limited value, though their prevalence makes them worthy of mention.

### 6.1.6 Industrially Relevant High Pressure Enzymology

Limited information is available on the effect of pressure on enzymes that are not model systems, such as those that have recently been isolated or have been shown to catalyse non-classical chemical reactions. To date, no investigation into the effects of HP on enzyme mediated catalysis have been performed upon enzymes involved in the biosynthesis of natural products. The complex chemistry involved in these pathways simultaneously drives intense interest, and often precludes the use of highly specialised equipment with its necessary limitations.

### 6.1.7 The Aim of this Work

The paucity of information pertaining to the effect of pressure on NP biosynthetic enzymes, in particular those isolated from marine microorganisms recovered from extreme pressures, represents a substantial knowledge gap. The isolation of *M. maris* and the subsequent characterisation of AbyU, coupled with the well reported effects of pressure on a synthetic Diels-Alder reaction,[388] makes the AbyU-catalysed DA reaction an excellent target for investigation. Such a study would help to inform our knowledge of the effect of pressure on a wider variety of enzymatic processes and architectures, as to date β-barrel enzymes have not yet been subjected to characterisation under HP regimes. For these reasons the purpose of the work reported in this chapter was to undertake a detailed molecular level characterisation of the effect of HP on AbyU. To achieve this, a combination of HP-SF and HP-SRCD were used. These techniques allow us to observe overall conformational change and stability, as well as the kinetic profile of the enzyme under investigation.

## 6.2 Methods

### 6.2.1 Cloning and Expression of AbyU

AbyU was cloned and expressed as previously described, with the exception that SEC purification was performed using a GE Healthcare 16/60 S75 column, pre-equilibrated in CD buffer (10 mM sodium phosphate, pH 7.5). Fractions were analysed by absorbance at 280 nm and SDS-PAGE. AbyU was found to be stable in this buffer. Purified protein was concentrated to 500 μM and flash frozen in liquid nitrogen or stored at 4 °C for a maximum of 14 days.

### 6.2.2 The High Pressure SRCD System

A high-pressure cell (HP-200) (ISS), designed for fluorescence and absorption studies, was used for all HP-SRCD experiments. The cell was interfaced to beamline B23, at Diamond Light Source, Oxfordshire UK. The cell is made from a stainless-steel alloy and can withstand pressure of up to 2000 bar with $MgF_2$ windows (ISS). The windows, 10 mm in diameter and 8.5 mm thick, are mounted into the cell via a Poulter type configuration with 10 mm clear aperture. During the experiment, $N_2$ was blown over the windows to clear any water leakage from the windows. Aqueous samples were loaded into custom made circular quartz cells (ISS) of diameter 8 mm and stoppered with a rubber cap, expelling all air. The cuvette was placed in a holder and then positioned into the HP cell. The interior of the cell was subsequently filled so that the sample cell was completely surrounded by the pressurising medium, clean, degassed, de-ionised water. The pressure cell is mounted onto the beamline within a blackout box to prevent any light from interfering with the signal. The pressure cell is connected, *via* flexible tubing, to a manually operated hydrostatic pressure pump. Two valves are included in the pressure system which allow the pump to be isolated from the reservoir and or the high-pressure cell. The reservoir, pump and tubing are filled with degassed and deionized water and cleared of any air bubbles prior to use.

### 6.2.3 Circular Dichroism Spectroscopy of AbyU at High Pressure

Far-UV CD spectra were recorded using an AbyU protein solution of 1.5 μM in CD buffer supplemented with 10 % acetonitrile, 10 % methanol or 40 % methanol. Baseline readings

were collected from CD buffer supplemented with the relevant organic solvent at a range of pressures up to 2000 bar. Baselines were subtracted from the relevant AbyU CD spectra obtained at the same pressures. Samples containing AbyU were equilibrated for 120 seconds at pressure before readings were taken. Four repeats of CD spectrum were collected and averaged between 190 – 260 nm. Data were initially processed by the subtraction of the relevant baseline, and the removal of results where the high-tension voltage (HT) exceeded 700 V. Results were plotted using Graphpad Prism. Secondary structure analysis was performed using BestSel.[405]

### 6.2.4 High Pressure Stopped-Flow Experiments

HP-SF experiments were performed using a Hi-Tech Scientific HPSF-56 high-pressure stopped-flow spectrophotometer (TgK Scientific, Bradford on Avon, UK). The stopped-flow spectrometer was zeroed and references taken using dd$H_2O$. For data collection, Syringe 1 contained a final concentration of 40 μM AbyU, in 10 mM phosphate buffer. Syringe 2 contained substrate **9** diluted into 10 mM phosphate buffer with 20% acetonitrile. Syringes were placed within the pressure cell and equilibrated briefly at increasing pressures before 50 μL from each syringe was injected into the mixing chamber and measurement immediately commenced. Data were measured at 325 nm using a logarithmic timebase for 5 seconds. All reactions were performed at 22 $^{o}$C, and 3 – 5 transients were averaged for all data points. Results were fit using Graphpad Prism. Initial concentrations of substrate **9** were calculated from the absorbance at t = 0, to account for uncatalysed substrate turnover, as well as the tendency of the substrate to stick within the apparatus. Data obtained before 2 ms was discarded. The data were subsequently fit to Equation 7. Measured amplitude changes in absorbance were converted to concentration using the measured absorbance coefficients for the substrate and product described previously. Data for the rate of background turnover of **9** were also recorded at 325 nm. 10 mM Sodium Phosphate buffer in syringe 1 was mixed with 50 mM Substrate **9** in 10 mM Phosphate buffer supplemented with 20% acetonitrile, and the data were fit to a single exponential (Equation 15). The first 0.5 seconds of each run were discarded due to mixing effects generated by the presence of acetonitrile in syringe 2 but not syringe 1.

Equation 15:

$$\mathrm{Y} = \mathrm{Y} = \mathrm{A}_0 + \mathrm{A}_1 \times \exp(-k_1 \times \mathrm{X})$$

## 6.3 Results: Structural Stability of AbyU at High Pressure

Given that AbyU was isolated from a deep-sea marine microorganism, it was hypothesised that it could have developed to function at HP. This hypothesis is particularly appealing as HP environments are known to drive DA reactions synthetically. This may be of relevance in the context of employing AbyU in high pressure industrial processes, with related investigations into the compatibility of this DA biocatalyst in the presence of high organic solvent concentrations recently undertaken.[351] Studies conducted under HP regimes may also serve to provide further insight into the kinetic behaviour of AbyU. In order to probe this hypothesis, a combination of structural and transient kinetic studies, using both HP-SF and HP-SRCD, were performed at a range of HP regimes and in the presence of a variety of solvents.

### 6.3.1 Probing Structural Changes in AbyU Induced by High Pressure in the Presence of Organic Solvents

In order to gain insight into pressure induced structural changes in AbyU, it was necessary to monitor any the behaviour of this enzyme in response to high pressure *in situ*, as this gives considerably more insight than simply determining the point at which AbyU is denatured by pressure. This necessitated the use of a state-of-the-art HP-SRCD system. HP-SRCD is a novel technique, which to date has only been used once in the study of enzyme structure.[401] This technique uses the extra intensity of a synchrotron radiation to penetrate the water used as a pressurising medium, facilitated by the use of custom $MgF_2$ windows of 8.5 mm thickness.

AbyU was incubated within the HP-SRCD cell with either 10 % acetonitrile, 10 % methanol or 40 % methanol. These solvents are often used within industrial process, and are required to ensure the solubility of substrate **9**, a synthetic analogue of AbyU's natural substrate. These conditions are, therefore, far more relevant to study from a practical point of view, although the solvents do prevent high-quality data from below 200 nm from being obtained. HP-SRCD data reports on the overall structural integrity of the target protein, in this case AbyU, and reveals changes in secondary structure composition. These changes can be monitored at any discrete pressure up to 2000 bar (200 MPa). The sample is then returned to atmospheric pressure, whilst monitoring any changes in secondary structure brought about by this process.

Initially, structural changes in AbyU at HP were monitored in the presence of 10 % acetonitrile (Figure 6.1). This revealed that AbyU retains an intrinsically folded conformation at all pressures investigated up to 2000 bar. The HP-SRCD transients were subsequently interpreted as a function of secondary structure composition, as calculated using the program BestSel.[405] This program uses a comparison between secondary structure elements observed by X-ray crystallography and the CD spectrum of the equivalent parent polypeptide to determine a prediction of the secondary structure composition from a CD spectrum. This analysis is consistent with AbyU being principally composed of a β-strands. At increased pressures the small proportion of α-helix present in the structure appears to be lost at HP, whilst the proportion of loops and flexible regions decreases slightly. This contrasts with a significant increase in the proportion of β-strand content, which rises from ~ 40 % to > 50 %. This additional β-sheet structure appears to remain once pressure is released, with the initial α-helical content not recovered.

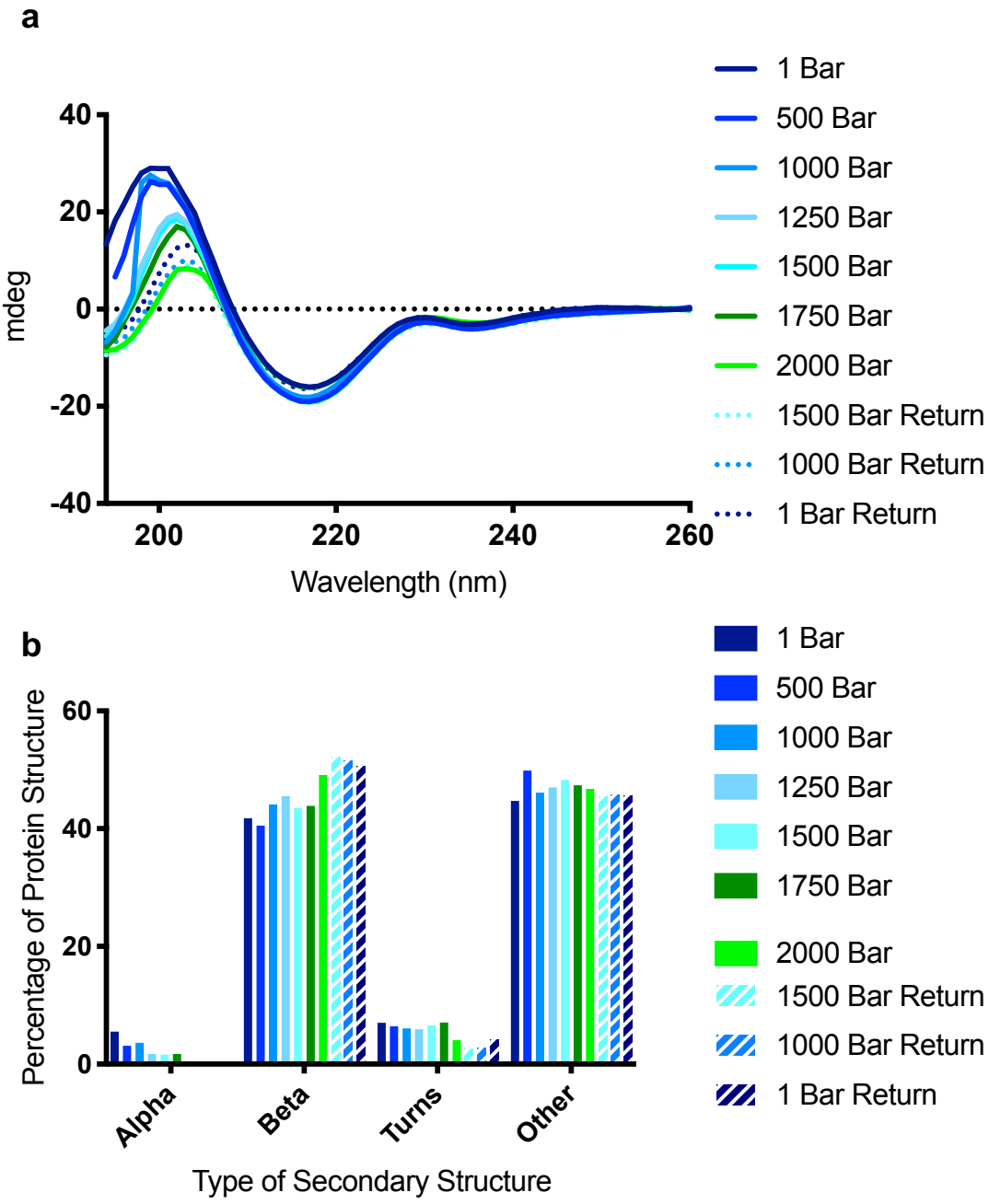



**Figure 6.1 HP-SRCD of AbyU in the Presence of 10 % Acetonitrile.** a) HP-SRCD spectra of AbyU at pressures between 1 and 2000 bar. HP-SRCD spectra were also obtained whilst returning the sample to atmospheric pressure from 2000 bar. These data are labelled 1500 bar return, 1000 bar return and 1 bar return respectively, and are indicated by dotted lines. b) A prediction of the secondary structure composition of AbyU at pressures between 1 and 2000 Bar. The prediction was performed by BestSel using the HP-SRCD spectrum between 200-250 nm for AbyU at the corresponding pressure. Secondary structure predictions labelled 'Return' indicate the prediction for the secondary structure

once the sample has had the pressure decreased to that value from 2000 Bar, and are indicated by dashed bars.

In order to further investigate the structural stability of AbyU, a harsher solvation medium was used. Although increasing the concentration of acetonitrile present in the medium was initially explored, this did not allow reliable HP-SRCD results to be obtained, as the solvent interferes with the passage of polarised light through the cuvette. As methanol interferes less strongly with polarised light, and because AbyU has already been shown to be stable in its presence, this solvent was also trialled. Initially, 10 % methanol was used at pressures up to 1000 bar (Figure 6.2). In order to conserve the $MgF_2$ window, which becomes damaged by extensive exposure to the highest pressure regimens, HP-SRCD data was not obtained for pressures between 1000 – 2000 bar. Following data analysis using BestSel, a small increase in β-strand content at pressures up to 750 bar was identified, although at 1000 bar, the β content of the polypeptide was seen to decrease slightly. The results obtained during a return to atmospheric pressure indicate that this drop is likely to be within the constraints of experimental error.

The highest possible concentration of methanol that still allowed HP-SRCD data to be collected (40 %) was subsequently used in an attempt to destabilise the structural integrity observed at the highest pressures (Figure 6.3). This was ultimately unsuccessful, with AbyU retaining structural integrity at all pressures up to 2000 bar. An increased level of noise is observed in the spectra due to the higher concentration of methanol. Nevertheless, following data deconvolution in BestSel, an increasing trend in the proportion of β-content is observed. The higher solvent content also appears to have disrupted the small proportion of helical content, although this may be due to the higher level of noise in the spectra.

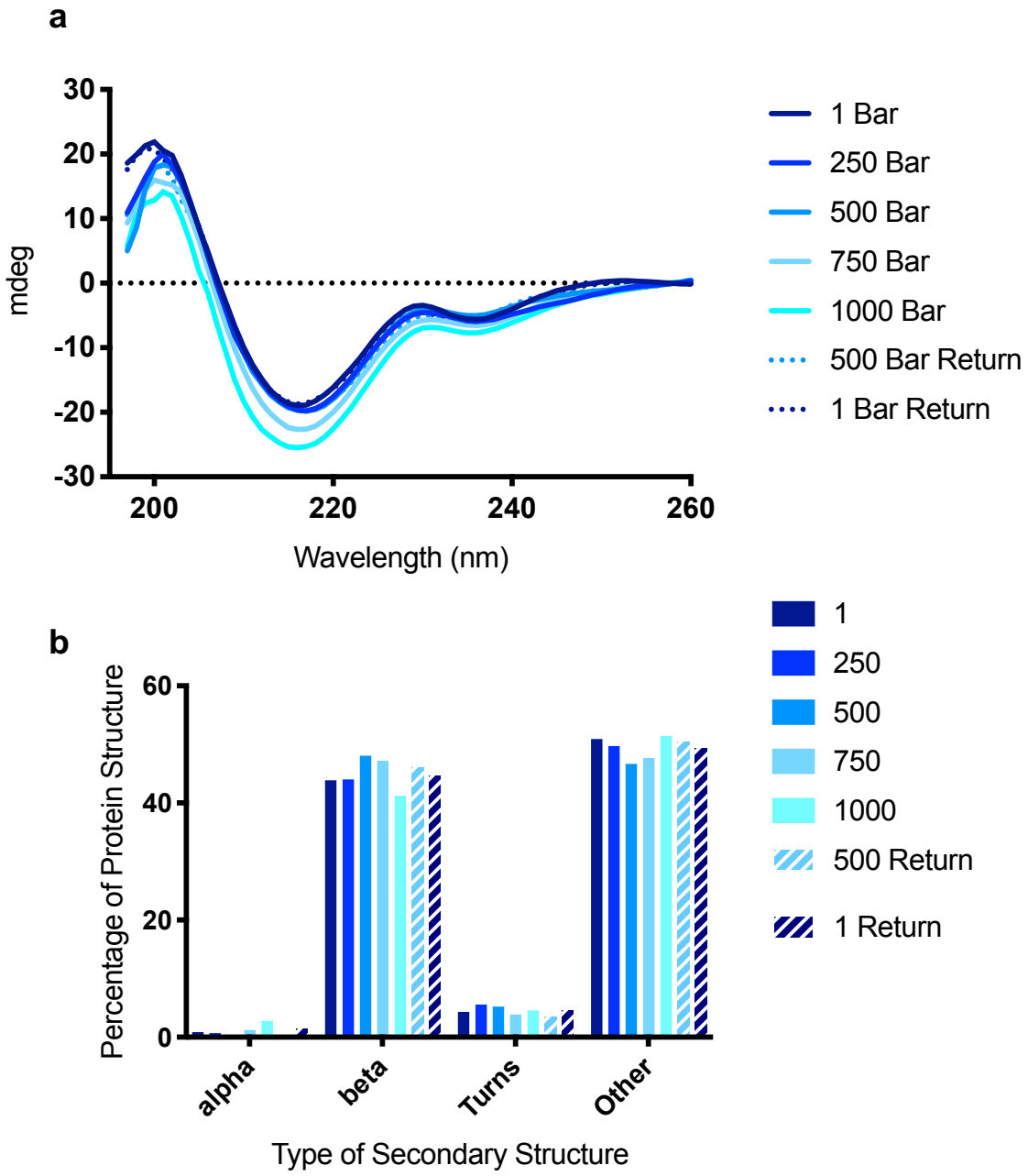



**Figure 6.2 HP-SRCD of AbyU in the Presence of 10 % Methanol.** a) HP-SRCD spectra of AbyU at pressures between 1 and 1000 bar. HP-SRCD spectra were also obtained whilst returning the sample to atmospheric pressure from 2000 bar. These data are labelled 1500 bar return, 1000 bar return and 1 bar return respectively, and are indicated by dotted lines. b) A prediction of the secondary structure composition of AbyU at pressures between 1 and 1000 Bar. The prediction was performed by BestSel using the HP-SRCD spectrum between 200-250 nm for AbyU at the corresponding pressure. Secondary structure predictions labelled 'Return' indicate the prediction for the secondary structure once the sample has had the pressure decreased to that value from 1000 Bar, and are indicated by dashed bars.

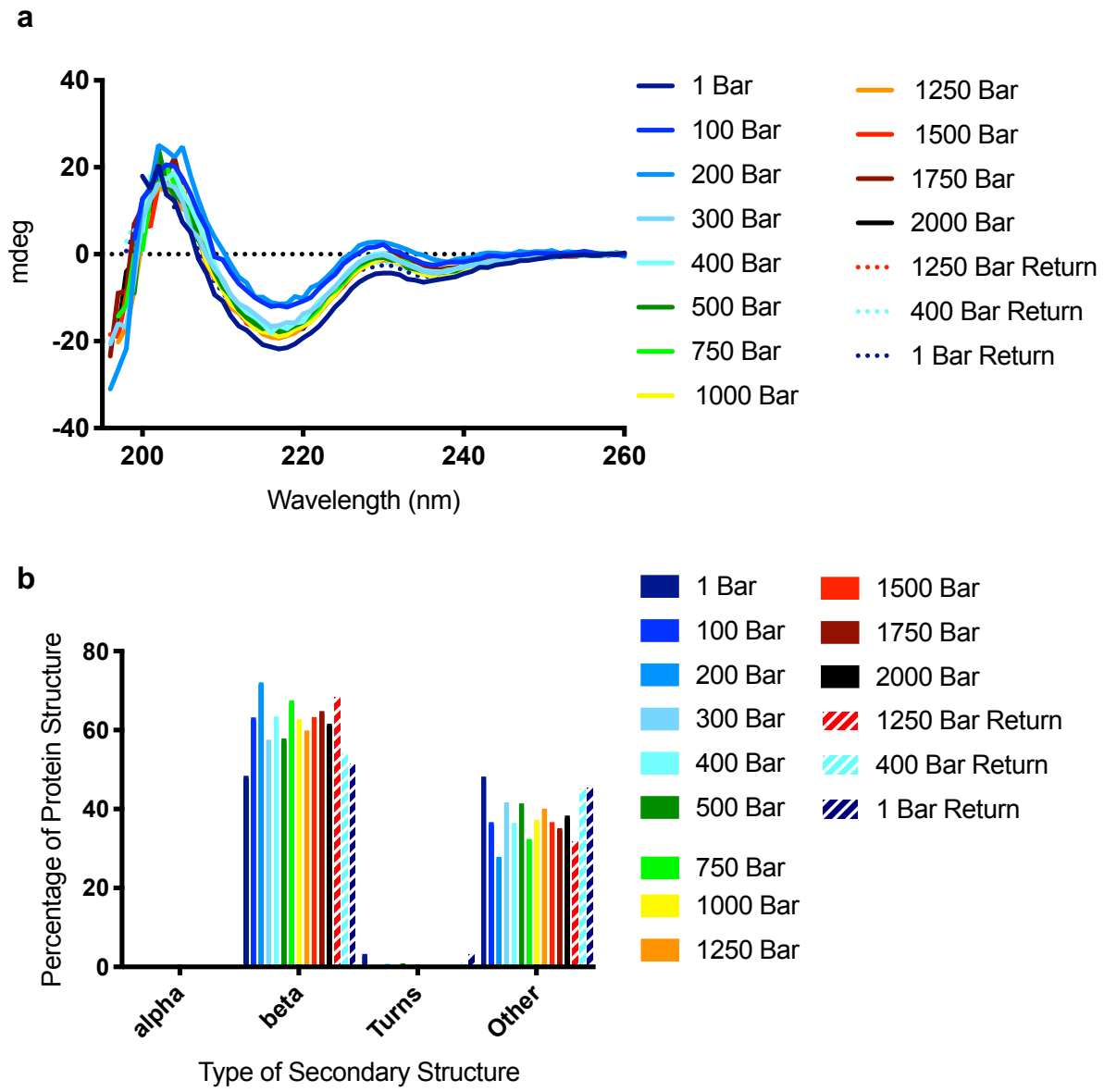



**Figure 6.3 HP-SRCD of AbyU in the Presence of 40 % Methanol.** a) HP-SRCD spectra of AbyU at pressures between 1 and 2000 bar. HP-SRCD spectra were also obtained whilst returning the sample to atmospheric pressure from 2000 bar. These data are labelled 1500 bar return, 1000 bar return and 1 bar return respectively, and are indicated by dotted lines. b) A prediction of the secondary structure composition of AbyU at pressures between 1 and 2000 Bar. The prediction was performed by BestSel using the HP-SRCD spectrum between 200-250 nm for AbyU at the corresponding pressure. Secondary structure predictions labelled 'Return' indicate the prediction for the secondary structure once the sample has had the pressure decreased to that value from 2000 Bar, and are indicated by dashed bars.

Together, these data demonstrate that AbyU maintains overall structural integrity at pressures up to 2000 bar and in the presence of all the solvents tested. Although no major global structural changes were observed, the HP regime investigated did raise the percentage of β-content significantly; by an average of 32 %, between 1 and 1000 bar (Figure 6.4). The proportion of β-strand then stabilises at around 60 %, though small increases may still be occurring up to 2000 bar. Some retention of elevated β-structure is observed once the sample is depressurised. This elevation is likely due to the HP environment strengthening existing hydrogen bonds and potentially driving the formation of new ones.[387] The backbone hydrogen bonds between each β-strand in the barrel of AbyU are therefore strengthened, and residues at either end of each strand may be incorporated.

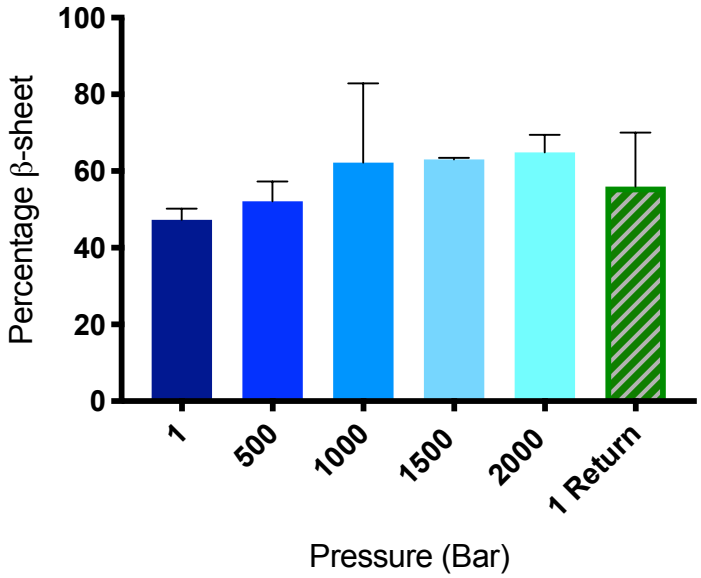



**Figure 6.4 The Influence of Pressure on the Proportion of β-Sheet Structure Observed in AbyU.** The average up to 1000 bar consists of data obtained in the presence of 10 % and 40 % methanol as well as 10 % acetonitrile. The average from 1500 bar to 2000 bar consists of data obtained in the presence of 10 % acetonitrile and 40 % methanol. Error bars are standard error from the mean.

## 6.4 Results: AbyU Reaction Kinetics at High Pressure

Whilst CD data are generally considered less informative for the study of predominantly β-structures, these data do show that attempts to disrupt the structural integrity of AbyU *via* a HP regime, serves to increase the proportion of β-content and decrease the proportion of loops and turns. Higher concentrations of solvent may also contribute to this effect.

In tandem with studies of the impact of HP on AbyU structure, it was also important to explore the impact of different HP regime on the kinetic performance of this enzyme. Due to the well-documented relationship between protein folding states and functional states,[406] it was hoped that the relationship between the HP-SRCD data and HP kinetic data would allow further mechanistic insight to be derived. The data would also enable assessment of the potential for using AbyU as an industrial catalyst at HP. As shown previously, AbyU possesses a complex reaction mechanism. Substrate **9** can be bound in at least three different conformations, with small reorganisations within the active site often necessary for catalysis. A number of practical complexities also prevented easy characterisation of the kinetic parameters of AbyU with respect to pressure. These included the propensity for **9** to cyclise spontaneously, and the need to incubate enzyme and substrate together for between 2 – 5 minutes in a static cell before the correct pressure regime was reached. The mechanistic and practical complexities together prevented the use of steady-state kinetics to investigate the effect of pressure on AbyU-catalysed cyclisation of **9** to form **10**. In light of these issues experiments were chosen to be performed using a HP-SF system, which removes the need to incubate the substrate and enzyme together prior to data collection. It also allows closer inspection of the effect of HP on the mechanism of AbyU. Traditional approaches to determining the effect of pressure on the catalytic efficiency of an enzyme, such as taking measurements before and after the application of a HP regime, were considered to be of limited value when assessing AbyU's catalytic performance under HP.

### 6.4.1 High-Pressure Stopped-Flow Transients of AbyU-Catalysed Cyclisation

Although the HP-SF characterisation of AbyU is only interpretable once the mechanism employed during catalysis has been deciphered, it was these experiments that first revealed the presence of multiple exponential phases in the AbyU catalysed reaction. As such, and due to a

lack of immediate access to HP-SF equipment, some of the experimental parameters used are now considered sub-optimal. They do, however, still provide important insight into the effect of pressure and the increase in β-strand structure on the catalytic mechanism of AbyU.

HP-SF experiments were performed by mixing 20 μM AbyU rapidly with **9**, with *in-situ* burst phase kinetic data collected at pressures between 1 bar and 1000 bar. The kinetic transients generated bare good resemblance to those obtained at atmospheric pressure (Section 5.3.7). These were analysed and fit to Equation 7 (Figure 6.5). Although a fit consisting of three exponentials and a slope was previously deemed an overfit in the presence of 1 μM AbyU, the increased concentration of enzyme used in these experiments induced the manifestation of a third burst phase. As the size of the burst phase is approximately stoichiometric with the concentration of AbyU in the reaction, this is likely to represent a third binding conformation. The manifestation of this additional phase in $k_{b2}$, the slower of the two exponentials, indicates that it may represent the second, weaker, binding conformation identified in $k_{obs2}$ during single turnover analysis.

Equation 7:

$$\mathrm{Y} = \mathrm{A}_0 + \mathrm{A}_1 \times \exp(-k_1 \times \mathrm{X}) + A_2 \times \exp(-k_2 \times \mathrm{X}) + A_3 \times \exp(-k_3 \times \mathrm{X}) + (\mathrm{Slope} \times \mathrm{X})$$

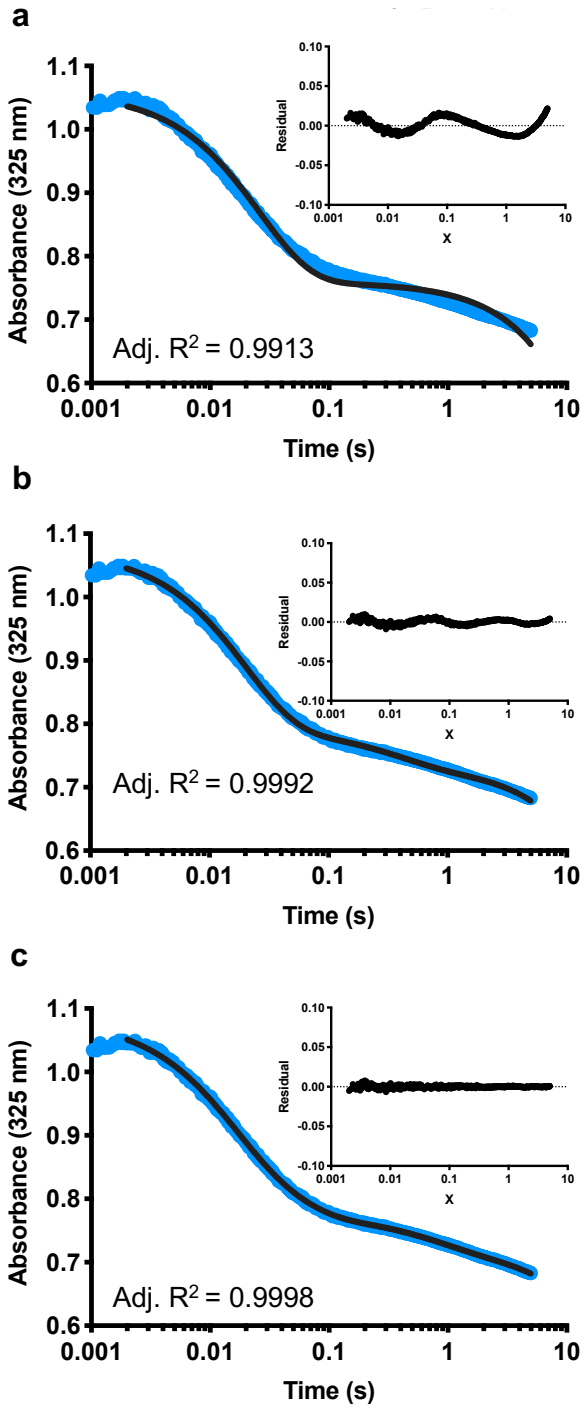



**Figure 6.5 An Example Transient and Fitting Statistics for AbyU-Catalysed Cyclisation at HP.** This transient was produced by the turnover of 83 μM substrate **4** by 20 μM AbyU at 750 bar. a) Fit to Equation 6. b) Fit to Equation 5. c) Fit to Equation 7. Adjusted $R^2$ values are shown beneath the transients. The residual plots generated are shown inset for comparison.

### 6.4.2 Spontaneous Cyclisation of 9 at High Pressure

The next goal was to investigate the effect of pressure on the rate of spontaneous cyclisation of **9** to **10** (Figure 6.6). Similarly to other Diels-Alder reactions, the cyclisation of **9** in the absence of enzyme is driven at an increasing rate by increasing pressures. Interestingly, the spontaneous turnover of **9** at HP occurs in an exponential manner, whilst the rate of reaction at atmospheric pressure remains linear. This indicates an increased level of cooperation between substrate molecules at HP. One possibility is that, at HP, the reaction volume is decreased, leading to an increased proliferation of the higher order structures observed by DLS (Section 5.2.1). Injection into the reaction chamber of the stopped-flow subsequently halves the concentration leading to the disassembly of the higher order structures and an increased initial rate of spontaneous cyclisation. The rate of spontaneous cyclisation of **9** increases by around 10 % of the initial rate for every 100 bar increase in pressure (Figure 6.6b). Due to the necessity of filling two large syringes with reactants and then progressively increasing the pressure on those reactants prior to injection, the quantity of starting material decreased at a rate proportional to the pressure and concentration of substrate **9** present in the syringe (Figure 6.7). This means that not only is it impossible to realistically obtain multiple repeats of data with the same starting concentration at different pressures, but that the concentration of starting material in the syringe can decrease by roughly a third over the course of the experiment.

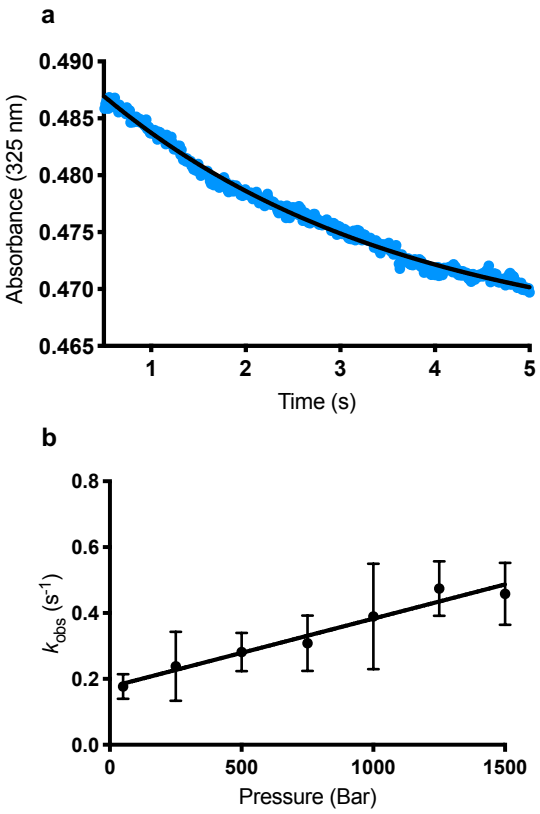



**Figure 6.6 Pressure Dependence of the Rate of Spontaneous Cyclisation of 9.** a) Example transient for the spontaneous turnover of Substrate **9** at 500 bar. Data were fit to a single exponential and the rate *k* derived for each replicate. b) The rate of spontaneous cyclisation as a function of pressure. Every 100 bar increase in pressure results in an increase in rate of 0.02 ± 0.004 $s^{-1}$.

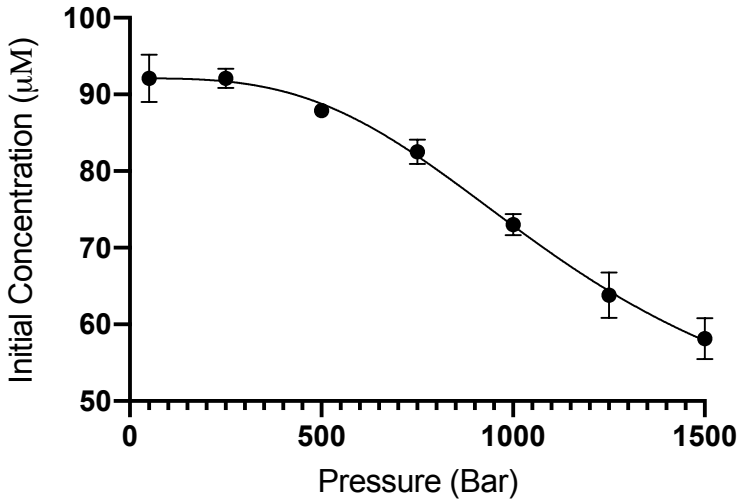



**Figure 6.7 The Initial Concentration of Substrate 9 at Increasing Pressures.** Each run was initiated from the same syringe, which is incubated at increasing pressures before injection. The concentration was calculated from the initial absorbance and the measured extinction coefficient, and then fit to Equation 11.

### 6.4.3 Determining a Concentration Dependence at Atmospheric Pressure

In order to account for the decrease in the concentration of **9** with increasing pressure, a full concentration dependent dataset for the reaction of 20 μM AbyU with substrate **9** was obtained at atmospheric pressure. The individual transients were analysed and fit to Equation 7 (Figure 6.8). The rate and amplitude of each phase within the reaction was determined and plotted with respect to the concentration of substrate (Figure 6.9). This allows us to extract the three discernible burst phases under these reaction conditions, determining their rates, $k_{b1} - k_{b3}$, and their amplitudes, $A_{b1} - A_{b3}$. The relationship between the concentration of substrate and the three phases of the reaction could then be determined. The parameters for the equations linking concentration, rate and amplitude are given in Table 6.1.

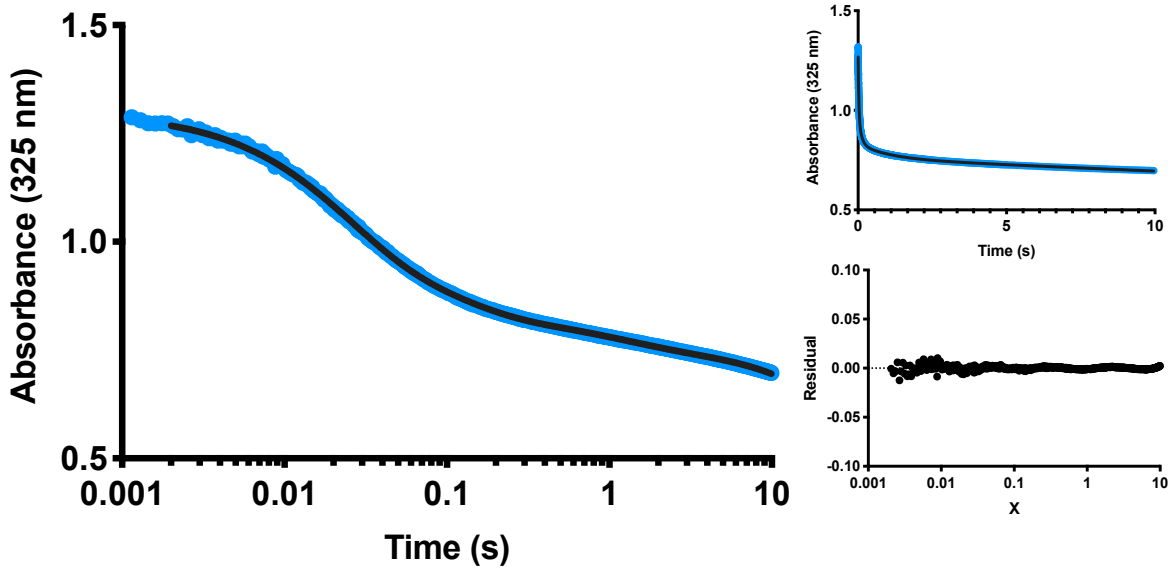



**Figure 6.8 An Example Transient for the Incubation of 102 µM Substrate 9 with 20 µM AbyU at Atmospheric Pressure.** The same transient is shown on both a logarithmic and linear scale for comparison, as is the residual trace for the fit to Equation 7.

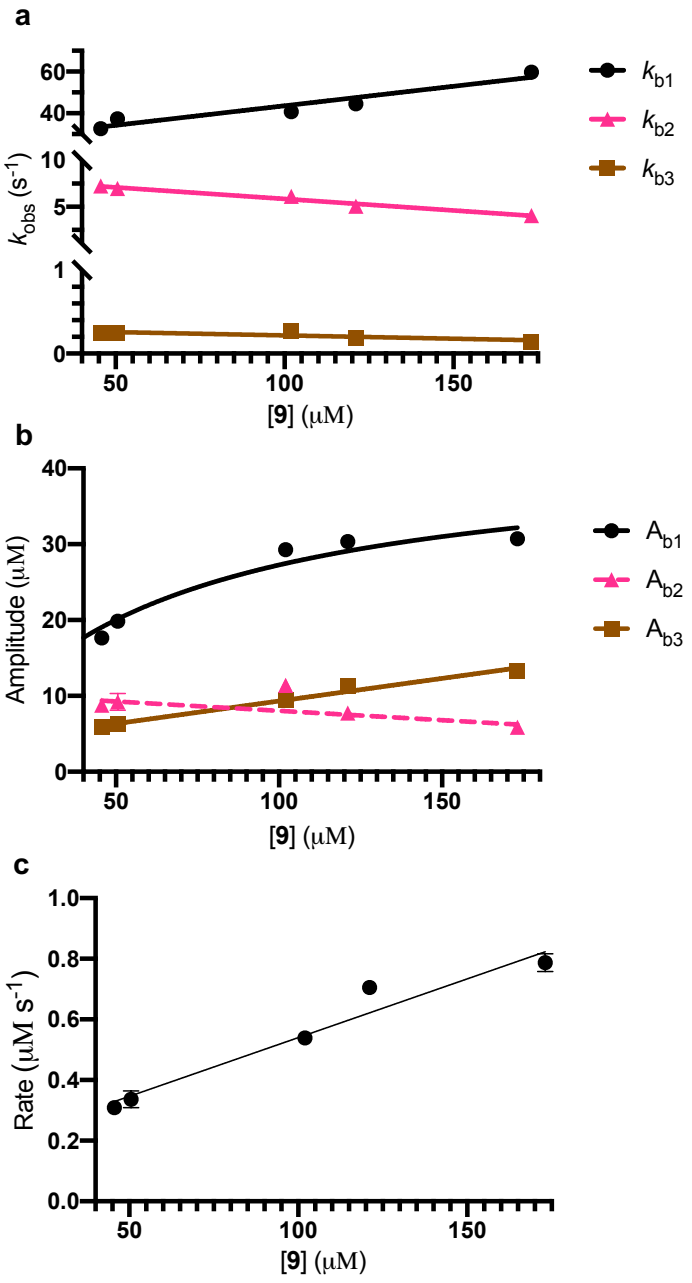



**Figure 6.9 Kinetic Data Derived from the Reaction of 20 μM AbyU with Varying Concentrations of Substrate.** a) The relationship between $k_{b1}$, $k_{b2}$ and $k_{b3}$ and the concentration of substrate present in the reaction. Each relationship is fit to a straight line (Equation 1). b) The amplitude of each of the three burst phases, $A_{b1}$, $A_{b2}$ and $A_{b3}$, and their change with respect to substrate concentration. Whilst $A_{b2}$ and $A_{b3}$ have a linear relationship to substrate concentration and are fit to Equation 1, $A_{b1}$ is fit to a rectangular hyperbola (Equation 10). c) The rate of the steady-state reaction with respect to substrate concentration. Under these non-saturating conditions, this is fit to a straight line.

| Phase | Slope ($\mu M^{-1}$) | Y-intercept ($\mu M$) |
|---|---|---|
| $k_{b1}$ | 0.19 ± 0.015 | 24.9 ± 1.7 |
| $k_{b2}$ | -0.025 ± 0.0020 | 8.34 ± 0.23 |
| $k_{b3}$ | -0.00079 ± 0.00014 | 0.30 ± 0.016 |
| Steady-State | 0.0039 ± 0.00022 | 0.15 ± 0.025 |
| $A_{b2}$ | -0.025 ± 0.0045 | 10.50 ± 0.52 |
| $A_{b3}$ | 0.060 ± 0.0020 | 3.35 ± 0.22 |
| | $A_{lim}$ ($\mu M$) | $K_d$ ($\mu M$) |
| $A_{b1}$ | 42.71 ± 1.8 | 56.66 ± 6.4 |

**Table 6.1 Parameters for the Concentration Dependence of a Reaction with 20 μM AbyU.**

### 6.4.4 The Effect of Pressure on Concentration Independent AbyU Catalysis

Data acquired from AbyU HP-SF experiments do not constitute a formal kinetic interrogation of AbyU, as the concentration of enzyme present is too high to ensure pseudo-first order kinetics. The data do, however, allow us to establish the effect that a decrease in substrate concentration has on the rate and amplitude of each phase observed in the HP experiments. The parameters given in Table 6.1 were used to account for the effect of the varying substrate concentration on each transient generated at HP. The expected change in rate or amplitude for each phase due to the decreasing concentration of substrate was determined. The HP transients were subsequently analysed with these expected changes taken into account. This allowed the determination of a concentration independent pressure effect on the cyclisation of substrate **9** by AbyU (Figure 6.12). The relationship between pressure and the rate and amplitude of each phase was determined. This provided further insight into the enzyme mechanism and also the limits placed by HP regimes on catalysis.

Unfortunately, it was not possible to conduct the same analysis for the enzyme independent reaction, due to the changing nature of the transient observed. It should therefore be noted that the background rate of reaction has not been corrected for the variation in initial substrate concentration. It is not anticipated that this will have had any bearing on the results obtained from the AbyU catalysed reactions due to the disparity in enzyme catalysed and uncatalysed rates.

The rate of reaction observed in each of the three burst phases increases with increasing pressure. This is possibly due to the smaller reaction volume present at HP causing an increase in the rate of collisions. This is likely to account for the small increases observed in $k_{b2}$ and $k_{b3}$, as the amplitudes corresponding to these phases remain unaffected by a change in pressure. $A_{b1}$, however, decreases by roughly 2/3 over the pressures analysed. This reduction in amplitude represents fewer molecules of **9** binding in this conformation and successfully cyclising before the rate-limiting product release step occurs. Given that $A_{b2}$ and $A_{b3}$ remain unaffected, and there are no major structural rearrangements occurring, we can infer that the increased rigidity of the AbyU β-barrel prevents small rearrangements of the substrate from taking place upon binding. This in turn leads to an increase in the number of substrate molecules binding in a non-productive conformation, such as D, as opposed to the most productive conformation, A. However, with rearrangement unable to occur, those molecules that do successfully bind in are able to cyclise significantly faster, with the rate of reaction in this phase increasing by 2-fold.

Although the reaction conditions do not saturate the steady-state rate of reaction, they still evidence the effect of HP regimes on AbyU. Up to ~500 bar, AbyU is still able to catalyse the cyclisation of **9** with the same efficiency. Beyond 750 bar, however, a drastic decrease in the rate of the steady-state reaction is observed, until, at 1250 bar, the rate of enzyme-catalysed reaction does not exceed the background rate of reaction. Given that the rate of reaction has been adjusted for the decrease in initial substrate concentration, there are two more possible hypotheses for this catalytic inhibition. The first of these is that the enzyme active sites are each rapidly inhibited by non-productively bound substrate. The sigmoidal shape of the curve, however, suggests that the pressure at which catalysis can no longer occur effectively is reached suddenly, and is not a linear process. When fit to Equation 11, the hill coefficient of -5.4 ± 2.4 (which increases to -8.8 ± 2.7 if the fit is constrained to 0 on the bottom) indicates a very high degree of cooperativity between pressure and the rate of reaction. A $K_{0.5}$ of 917 ± 110 provides us with an upper limit for the pressure regime AbyU can effectively function within. The concept that active sites might suddenly be unable to bind substrate in a productive conformation does not agree with the burst phase data, or the mechanistic understanding developed so far. Although non-productively bound substrate doubtless contributes to the decrease in rate, the burst phase data show that enzymatic catalysis does still occur, and that substrate is still able to bind in a productive conformation. What is more likely to be causing this highly cooperative

enzymatic inhibition is the rate-limiting product release step. If the capping loop is no longer dissociating from the active site to allow product release, the product would become trapped inside the β-barrel, creating a permanently inhibited EP complex. Another possibility is that, around 920 bar, the AbyU dimer dissociates, leading to an inactive population of monomers. There is no mechanistic or structural information to support the assumption that monomeric AbyU is inactive, however, so this theory should be discounted.

Equation 11:

$$Y = \frac{A_{lim} \times X^h}{{k_{half}}^h + X^h}$$

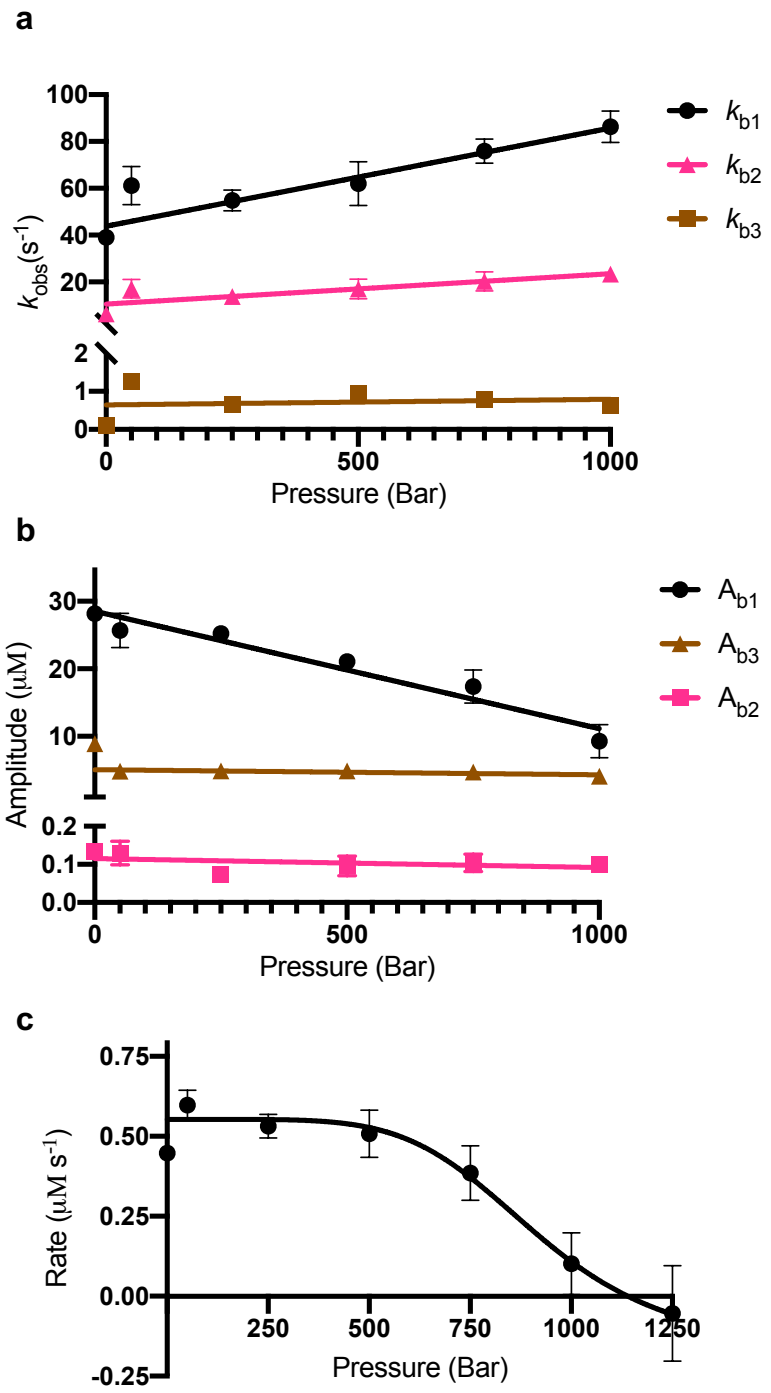



**Figure 6.10 The Concentration Independent Effect of Pressure on AbyU-Catalysed Cyclisation.** a) The rate of reaction in the three extractable burst phases in relation to pressure. The change in rate in relation to pressure is fitted to Equation 1 and the results shown in table 2. b) The effect of pressure upon the amplitude of each exponential phase in the reaction. The results are all fit to Equation 1 and the results are shown in table 2. c) The effect of pressure on the steady-state. The trend is fit to Equation 11. The rates of reaction shown above have all been normalised for concentration.

The parameters for the relationship between pressure and the AbyU catalysed cyclisation of **9** are given in Table 6.2. $k_{b3}$, along with $A_{b2}$ and $A_{b3}$, remain largely unaffected by the HP regimes investigated. $k_{b1}$ and $k_{b2}$ increase by around 10 % of their starting rate with every 100 bar increase in pressure. $A_{b1}$, however, decreases by roughly 6 % of its starting amplitude over 100 bar. The steady-state shows a high degree of cooperativity between pressure and catalysis, with AbyU unable to efficiently catalyse substrate conversion at pressures > 920 bar.

| Phase | Slope (μM *per* 100 bar) | Y-intercept (μM) |
|---|---|---|
| $k_{b1}$ | 4.2 ± 0.4 | 43.8 ± 2.5 |
| $k_{b2}$ | 1.3 ± 0.2 | 10.6 ± 1.3 |
| $k_{b3}$ | 0.015 ± 0.022 | 0.64 ± 0.13 |
| $A_{b1}$ | -1.74 ± 0.13 | 28.53 ± 0.7 |
| $A_{b2}$ | 0.00 ± 0.002 | 0.12 ± 0.009 |
| $A_{b3}$ | -0.08 ± 0.04 | 5.06 ± 0.2 |
| | $K_{0.5}$ (μM) | Hill Coefficient |
| Steady-State | 832 ± 36.8 | -8.2 ± 2.3 |

**Table 6.2 Parameters for the Pressure Dependence of AbyU-Catalysed Cyclisation of 9.**

## 6.5 Discussion

Whilst the experiments undertaken during this chapter do not represent a complete study of a model system and its function in a high pressure environment, they do provide a valuable insight into the effect of pressure on enzyme-mediated catalysis. They have also proven important in informing the design of subsequent experiments focused on exploring the effect of HP on enzyme systems.

The HP-SRCD data reported herein show that there are no major structural rearrangements within AbyU at pressures up to 2000 bar, in the presence of different solvents and concentrations of solvent. The stability exhibited by AbyU under these pressures is perhaps surprising given its sizeable active site cavity. Roche *et al*. demonstrate that staphylococcal nuclease, a protein containing a small β-barrel fold within its structure, begins to denature at pressures as low as 500 bar.[407] This study showed that the cavity within this β-barrel was the principal driver behind pressure dependent denaturation, due to its significant internal volume. Furthermore, the data also show that a moderate increase in pressure can promote β-strand formation within AbyU. Although HP-NMR would be required to determine whether this is the result of compaction or conformational transition to a higher energy state, HP-SRCD does give a wider, and more easily interpreted, overview of the structural state of AbyU at high pressure. The newly developed HP-SRCD technique exploited here provides the ability to observe the enzyme in its entirety, as opposed to the individual $^{15}N$ and $^{1}H$ shifts observed in HP-NMR. As such, these data are the first example of HP driving an overall increase in β-sheet architecture within an enzyme, and provide further evidence that HP can favour hydrogen bond formation within macromolecular structures.

The kinetic data obtained by HP-SF, when interpreted in combination with the structural data obtained by HP-SRCD, provide a unique insight into the reaction mechanism employed by AbyU. The increased rigidity of the β-barrel, brought about by increased numbers of shortened backbone hydrogen bonds, appears to prevent conformational rearrangement of the substrate within the active site. As discussed in Chapter 5, $k_{obs1}$ contains multiple slightly different binding conformations, some of which must undergo minor rearrangements until the substrate is poised for cyclisation in pose A, as identified by MD simulations (Figure 5.24). The decrease in amplitude observed in $A_{b1}$, points to a decrease in the number of minor rearrangements

occurring during the burst phase. Optimally oriented substrate is still able to bind, with the increased rate of reaction at high pressure reflecting a lower manifestation of less productive orientations. Binding in conformations B and D is not affected by the increase in pressure. Increased rates of reaction in these burst phases is likely due to the smaller reaction volume for the ES complex.

Steady-state kinetic studies reveal the effect of pressure on the rate-limiting product release step of the reaction. We see a highly cooperative relationship between pressure and the rate of reaction, indicating that, at ~920 bar, product release is no longer able to occur. As product release increases the reaction volume, this process will be naturally inhibited to some degree by increasing pressure. We would, however, expect this process to be linear in nature. The initial mechanistic interpretation is that the capping loop, which needs to dissociate from the top of the barrel to permit product release, is no longer able to open, preventing product release and trapping the EP complex. In this enzyme the substrate can enter the active site in a variety of conformations and at a variety of rates. In the same way that a lower temperature can give the substrate time to rearrange and accurately bind within the active site, a lower pressure can increase the flexibility of the active site. Whilst the rates of binding and turnover are slower under these conditions, this appears to allow the substrate the time required to bind productively.

This study showcases the mechanistic information that can be gained through the study of pressure perturbation on enzyme catalysts. We see that the robust β-barrel architecture is capable of withstanding pressurisation, despite its large internal cavity, and that the active site of AbyU can be locked shut under HP regimes. This information not only increases the basic understanding of the effects that pressure can have on macromolecules, but it also provides crucial information about protein folds that remain stable under adverse conditions and how they might be exploited for the creation of a HP industrial catalyst.

# Chapter 7

# Discussion

## 7.1 The Abyssomicin Biosynthetic Pathway

Although abyssomicin C is one of the smallest and most simple of the tetronate antibiotics,[198] the biosynthetic machinery used to make it consists of a multitude of enzymes capable of performing complex chemical catalysis (Figure 1.13). Following synthesis, the polyketide chain is liberated as a linear molecule, without the use of a TE domain, by a rare combination of three enzymes, AbyA1 – A3. This simultaneously creates the tetronate moiety that gives this class of natural products their name. The acetate-lyase AbyA5 performs a little seen chemical reaction, preparing the dienophile required for the DA cyclisation to occur in the next step.[225] The elimination of acetate without the production of hydrolysis product, along with the later enzymatic insertion of an epoxide before the creation of the active antimicrobial product, are two more unusual examples of enzymatic catalysis. This means that virtually every step in the biosynthesis of abyssomicin C represents a rarely characterised or not yet understood enzymatically catalysed reaction. This relatively short and simple biosynthetic pathway is an exemplary model of the catalytic and structural complexity that can be accessed thanks to advances in gene sequencing and synthesis.[51]

The *bona fide* natural DAase AbyU is an excellent example of this catalytic complexity.[224] Not only has it been shown to be a remarkably stable, industrially relevant, catalyst,[351] but it has also been revealed here as a deceptively complex catalyst with great potential for rational engineering, now that the catalytic complexities are more conclusively understood.

Although polyketide synthases, their mechanism and products have been studied extensively over the last two and a half decades,[99,108,141,142,151,170,258,266,408] a number of questions about the complex web of structural and functional relations that guides vectoral synthesis remain unanswered.[149] The type I bacterial mPKS genes present in the abyssomicin biosynthetic pathway have been shown to present a tractable alternative for investigation to the two prototypical bacterial polyketide synthase pathways, the erythromycin and pikromycin

pathways.[107,409] The abyssomicin polyketide synthase genes allow us to study a minimal module, from which basic structural and mechanistic understanding can be built up, as well as the opportunity to study an unusual module containing two inactive domains.

## 7.2 The Abyssomicin Polyketide Synthase Genes

### 7.2.1 Cloning and Heterologous Expression

Although successful expression of AbyB2 remained elusive, the cloning and assembly protocol devised provides an alternative and more modern method for working with these large and repetitive genes. The ability to rapidly assemble biosynthetic genes into multiple vectors for analysis in multiple expression strains is essential for future studies of the many unique natural products discussed, and for future attempts to engineer these modular enzymes in a directed manner.

The successful expression of AbyB3 does, however, highlight the strain that an *E. coli* host is put under through the successful expression of polyketide synthase modules. Minor variations in culture conditions proved to completely eliminate protein expression, whilst multiple unsuccessful expression trials for both AbyB2 and AbyB3 show a need to develop a more robust polyketide synthase production host in the future. With improved genetic techniques, the ability to rapidly shuttle genes between hosts and vectors should make this task far more achievable.

### 7.2.2 AbyB3 Structural Studies

The structural studies of AbyB3 go a long way towards reconciling the differing modular architectures observed in mammalian FASs, intact mPKS modules and mPKS KS-AT didomains. The crystal structure of AbyB3ΔACP reveals a very high level of sequence and structural homology between mPKS didomains found in different bacterial species. This further points towards one dominant modular architecture within this diverse class of enzymes and makes the reconciliation of the different postulated architectures a necessity. The intricate protein-protein interactions postulated as necessary for correct mechanistic function, for

example in the binding site of the ACP to the KS and AT,[114,142] depend entirely on the structural model used for study. This in turn biases any attempt to re-engineer these modular assembly lines and represents a huge bottleneck in our ability to develop them for the production of designer polyketides.

This study, although not yet concluded, points towards a significant rearrangement of the modular polyketide structure upon the addition of the ACP domains. Although the AbyB3ΔACP structure is highly similar to those observed in DEBS, the AbyB3 2D class averages, and, indeed, the micrographs themselves, bare a high resemblance to the 2D class averages obtained during the reconstruction of a cryo-EM map for the arched PikAIII.[141] Crucially, no particles are observed close to the size of the 150 Å long didomain structure solved by x-ray crystallography. The smaller, ~75 Å, arch shaped particles are good evidence that a significant conformational rearrangement has been precipitated by the addition of the ACP domains, and points to a major, and as yet unidentified, role in structural fidelity for mPKS ACP domains. The structurally conserved *post*-AT linker region could play a significant and unidentified role in the transmission of structural rearrangements throughout the module, ensuring vectoral synthesis. Interactions of the *post*-AT linker with the KS-AT linker could also enable the structural rearrangements necessary for the creation of the enclosed reaction chamber created by the catalytic arch conformation. Hydrophobic contacts observed to stabilise the linear conformation of the KS and AT domains in the didomain structures,[114] and in FAS structures,[152] as well as the hydrogen bonding observed between the *post*-AT linker, the KS domain and the AT domain and their linker,[114] could all be disrupted by the movement of the *post*-AT linker, for instance upon the addition of a downstream domain.

## 7.3 AbyU

### 7.3.1 The Catalytic Mechanism of AbyU

The DA reaction holds great potential as a tool in synthetic chemistry, facilitating the creation of cyclohexenes with up to four new stereocentres and two new carbon-carbon bonds in a single step, whilst creating no waste products. It is, however, difficult to catalyse and control and remains sparsely used in synthesis. Although, until recently, there were no enzymes characterised that were capable of performing a standalone [4 + 2] cycloaddition reaction, the last

five years has seen a number of them identified. Whilst VstJ, PyrE3, PyrI4, AbyU, SpnF, MalC and PhqE have all been characterised to different levels, mechanistic understanding for all but SpnF remains non-existent.[224,331,337,338,344] Whilst the mechanism of SpnF has been investigated in more detail, mostly through theoretical studies,[340–343] the possible [6 + 4] reaction mechanism makes it unique amongst those [4 + 2] cyclases characterised to date, and hence a bad model system. Although PyrI4 is the only other [4 + 2] cyclase to possess a β-barrel fold, each of the enzymes characterised to date have been identified to enable catalysis through the steric restriction of their substrate. The simplicity of mechanism that this statement engenders belies the highly complex reaction mechanism that has been shown to exist beneath the surface.

The work described herein evidences a number of complex interactions between the substrate and the hydrophobic core of the active site. Multiple subtly different binding conformations can be grouped into three or four distinct substrate orientations within the active site, each with different binding energies and energy barriers towards catalysis. Rearrangements within the active site are sometimes necessary to facilitate catalysis, whilst mis-binding can be seen to create a catalytically incompetent enzyme substrate complex. Whilst these mechanistic observations might directly apply to PyrI4, they also show that the structural and steady-state kinetic characterisations performed so far on putative DAases are likely to hide significant complexities. These should be determined to enable engineering attempts and directed mutagenesis to proceed in a rational manner.

### 7.3.2 The Influence of Pressure and Temperature on AbyU-Catalysed Cycloaddition

The catalytic mechanism of AbyU, as well as its ability to function in a hostile industrial environment, was further probed at high pressures and cold temperatures. High pressure increases the rigidity of AbyU's β-barrel structure. The proportion of β-sheet secondary architecture is increased as the backbone hydrogen bonds connecting the antiparallel strands are strengthened. Mechanistically, this both prevents substrate rearrangements and adjustments from occurring within the active site and increases the rate of collisions, decreasing the time AbyU has to ensure correct substrate binding. This results in a decrease the amplitude observed in the most catalytically competent phase of the reaction. At pressures above 600 bar, there is a rapid and cooperative decrease in the rate of the steady-state reaction. Given that there is no corresponding decrease in the burst phase reactions, this must be a result of the effect of

pressure upon the rate limiting product release step. The obvious mechanistic conclusion, which is now supported by MD simulations, is that the active site capping loop becomes locked shut at high pressures, underlining the importance of this loop to the overall catalytic cycle.

Initial results from experiments conducted at 5 °C appear to corroborate these findings. The low temperatures will decrease the rate of collisions and decrease the rate of substrate binding within the active site. This allows AbyU the necessary time to ensure correct substrate binding, allowing cyclisation to reach 100 % over 1000 s.

## 7.4 Future Work

### 7.4.1 Polyketide Structural Studies

The first priority should be to screen quantifoil grids with a decreased proportion of the grid dedicated to carbon support, such as R 2/1 grids. As the concentration of sample and correct freezing conditions have already been determined, it should be possible to rapidly collect a full sized cryo-EM dataset of AbyB3. Once processed an assessment can be made as to whether any additional information is required or an attempt to bias additional particle orientations must be made. Given the results presented here, a high-resolution structure of AbyB3 is within reach and should be obtainable within a reasonable timeframe. If the theories presented here are confirmed, this would enable a detailed study into the modular architecture of mPKS enzymes as well as the complex protein-protein interactions that facilitate vectoral synthesis.

Once that is complete, further efforts can be made to study the multimodular mPKS enzymes AbyB1 and AbyB2. Efforts to express these constructs in heterologous hosts closer in identity to *M. maris*, such as *S. coelicolor*, could be made. Additionally, the two modules present in AbyB2 could be separated for expression, and then re-joined *in vitro* for biochemical and structural analysis. This would enable the structures of polyketide synthase modules containing all three commonly occurring reducing enzymes, as well as one containing inactive KR and DH domains, to be determined.

### 7.4.2 AbyU

Future work on AbyU can build on the detailed mechanistic understanding developed here. Attempts to engineer AbyU as a multifunctional DA biocatalyst can now proceed in a directed manner. One of the first possible approaches should be to create small mutations in the loop region, aiming to increase the rate of catalysis by effecting the rate limiting step. These mutations may enable us to develop a better understanding of the rearrangements necessary to ensure the substrate binds in the correct conformation, further building a detailed mechanistic understanding of this deceptively complex enzyme. *In silico* modelling should be exploited to speed and direct experimental work, with the MD and QM/MM studies performed to date providing a good basis from which to build.

## 7.5 Conclusion

Two key enzymes within the abyssomicin C biosynthetic pathway have been further characterised using structural and mechanistic studies. This work provides a strong basis from which attempts to engineer enzymes within this pathway and analogous enzymes from other systems can proceed. The structural studies of AbyB3 allow us to reconcile the apparent structural differences seen in different mPKS modules. It enables the creation of one mPKS architectural blueprint for use in future work developing and rationally engineering polyketide synthase modules for the creation of designer polyketides. The mechanistic understanding of elusive [4 + 2] cyclases has been increased, with the transient methods described within the first example of their application to this class of enzyme. Understanding the complexities involved allow us to direct future attempts to develop industrial DA biocatalysts, as well as attempts to understand analogous systems.

# Bibliography


1. Fleming, A. On the antibacterial action of cultures of a penicillium, with special reference to their use in the isolation of B. influenzae. 1929. *B World Health Organ* **79**, 780–90 (2001).

2. Landers, T. F., Cohen, B., Wittum, T. E. & Larson, E. L. A Review of Antibiotic Use in Food Animals: Perspective, Policy, and Potential. *Public Health Rep* **127**, 4–22 (2012).

3. Brauer, R. *et al.* Prevalence of antibiotic use: a comparison across various European health care data sources. *Pharmacoepidem Dr S* **25**, 11–20 (2016).

4. Ivanovska, V., Muller, A., Schweickert, B., Sharma, P. & Tao, W. *WHO report on surveillance of antibiotic consumption: 2016-2018 early implementation*. (2018).

5. Adedeji, W. A. THE TREASURE CALLED ANTIBIOTICS. *Annals of Ibadan postgraduate medicine* **14**, 56–57 (2016).

6. Davies, J. & Davies, D. Origins and evolution of antibiotic resistance. *Microbiol Mol Biology Rev Mmbr* **74**, 417–33 (2010).

7. Buhl, M. *et al.* Molecular Evolution of Extensively Drug-Resistant (XDR) Pseudomonas aeruginosa Strains From Patients and Hospital Environment in a Prolonged Outbreak. *Front Microbiol* **10**, 1742 (2019).

8. Norrby, R. *ECDC/EMEA JOINT TECHNICAL REPORT: The bacterial challenge: time to react*. (2009).

9. O'Niell, J. *Antimicrobial Resistance: Tackling a crisis for the health and wealth of nations*. (2014).

10. Kapoor, G., Saigal, S. & Elongavan, A. Action and resistance mechanisms of antibiotics: A guide for clinicians. *J Anaesthesiol Clin Pharmacol* **33**, 300 (2017).

11. Bolintineanu, D., Hazrati, E., Davis, H. T., Lehrer, R. I. & Kaznessis, Y. N. Antimicrobial mechanism of pore-forming protegrin peptides: 100 pores to kill E. coli. *Peptides* **31**, 1–8 (2009).

12. Barlow, M. & Hall, B. G. Phylogenetic Analysis Shows That the OXA β-Lactamase Genes Have Been on Plasmids for Millions of Years. *J Mol Evol* **55**, 314–321 (2002).

13. Goethem, M. W. V. *et al.* A reservoir of 'historical' antibiotic resistance genes in remote pristine Antarctic soils. *Microbiome* **6**, 40 (2018).

14. Munita, J. M. & Arias, C. A. Virulence Mechanisms of Bacterial Pathogens, Fifth Edition. *Microbiology spectrum* **4**, 481–511 (2016).

15. ABRAHAM, E. P. & CHAIN, E. An Enzyme from Bacteria able to Destroy Penicillin. *Nature* **146**, 837–837 (1940).

16. Campbell, E. A. *et al.* Structural mechanism for rifampicin inhibition of bacterial rna polymerase. *Cell* **104**, 901–12 (2001).

17. Leclercq, R. Mechanisms of Resistance to Macrolides and Lincosamides: Nature of the Resistance Elements and Their Clinical Implications. *Clin Infect Dis* **34**, 482–492 (2002).

18. Donhofer, A. *et al.* Structural basis for TetM-mediated tetracycline resistance. *Proc National Acad Sci* **109**, 16900–16905 (2012).

19. Hiramatsu, K. *et al.* Genomic Basis for Methicillin Resistance in Staphylococcus aureus. *Infect Chemother* **45**, 117 (2013).

20. Pagès, J.-M., James, C. E. & Winterhalter, M. The porin and the permeating antibiotic: a selective diffusion barrier in Gram-negative bacteria. *Nat Rev Microbiol* **6**, 893–903 (2008).

21. Poole, K. Efflux-mediated antimicrobial resistance. *J Antimicrob Chemother* **56**, 20–51 (2005).

22. Thomas, C. M. & Nielsen, K. M. Mechanisms of, and Barriers to, Horizontal Gene Transfer between Bacteria. *Nat Rev Microbiol* **3**, 711–721 (2005).

23. Duckett, S. Ernest Duchesne and the concept of fungal antibiotic therapy. *Lancet* **354**, 2068–2071 (1999).

24. Durand, G. A., Raoult, D. & Dubourg, G. Antibiotic discovery: History, methods and perspectives. *Int J Antimicrob Ag* **53**, 371–382 (2018).

25. Lewis, K. Platforms for antibiotic discovery. *Nat Rev Drug Discov* **12**, 371–87 (2013).

26. Lewis, K. Antibiotics: Recover the lost art of drug discovery. *Nature* **485**, 439–440 (2012).

27. Li, G. & Lou, H.-X. Strategies to diversify natural products for drug discovery. *Med Res Rev* **38**, 1255–1294 (2017).

28. Hanson, J. R. *Natural Products: The Secondary Metabolites*. (Royal Society of Chemistry, 2003).

29. All natural. *Nat Chem Biol* **3**, 351–351 (2007).

30. Thirumurugan, D. & Cholarajan, A. An Introductory Chapter: Secondary Metabolites. in (ed. Raja, S. S.) Ch. 1 (IntechOpen, 2018). doi:10.5772/intechopen.79766.

31. Serra, S., Fuganti, C. & Brenna, E. Biocatalytic preparation of natural flavours and fragrances. *Trends Biotechnol* **23**, 193–8 (2005).

32. Chauhan, B., Kumar, G., Kalam, N. & Ansari, S. H. Current concepts and prospects of herbal nutraceutical: A review. *J Adv Pharm Technology Res* **4**, 4–8 (2013).

33. Kelsey, N. A., Wilkins, H. M. & Linseman, D. A. Nutraceutical Antioxidants as Novel Neuroprotective Agents. *Molecules* **15**, 7792–7814 (2010).

34. Babizhayev, M. A. Biological activities of the natural imidazole-containing peptidomimetics n-acetylcarnosine, carcinine and l-carnosine in ophthalmic and skin care products. *Life Sci* **78**, 2343–2357 (2006).

35. Menn, J. J. Contemporary frontiers in chemical pesticide research. *J Agr Food Chem* **28**, 2–8 (1980).

36. Simmons, L. T., Andrianasolo, E., McPhail, K., Flatt, P. & Gerwick, W. H. Marine natural products as anticancer drugs. *Mol Cancer Ther* **4**, 333–42 (2005).

37. Sun, Y., Xun, K., Wang, Y. & Chen, X. A systematic review of the anticancer properties of berberine, a natural product from Chinese herbs. *Anti-cancer Drug* **20**, 757–769 (2009).

38. Kinghorn, A. D., Chin, Y.-W. & Swanson, S. M. Discovery of natural product anticancer agents from biodiverse organisms. *Curr Opin Drug Disc* **12**, 189–96 (2009).

39. Prakash, O., Kumar, A., Kumar, P. & Ajeet, A. Anticancer Potential of Plants and Natural Products: A Review. *Am J Pharmacol Sci* **1**, 104–115 (2013).

40. Karpiński, T. M. & Adamczak, A. Anticancer Activity of Bacterial Proteins and Peptides. *Pharm* **10**, 54 (2018).

41. Charlton, J. L. Antiviral Activity of Lignans. *J Nat Prod* **61**, 1447–1451 (1998).

42. Yokokawa, F., Asano, T. & Shioiri, T. Total Synthesis of the Antiviral Marine Natural Product (−)-Hennoxazole A. *Org Lett* **2**, 4169–4172 (2000).

43. Likhitwitayawuid, K. *et al.* Phenolics with antiviral activity from Millettia Erythrocalyx and Artocarpus Lakoocha. *Nat Prod Res* **19**, 177–182 (2005).

44. Bowling, J. J. *et al.* Antiviral and anticancer optimization studies of the DNA-binding marine natural product aaptamine. *Chem Biol Drug Des* **71**, 205–15 (2008).

45. Koehn, F. E., McConnell, O. J., Longley, R. E., Sennett, S. H. & Reed, J. K. Analogs of the Marine Immunosuppressant Microcolin A: Preparation and Biological Activity. *J Med Chem* **37**, 3181–3186 (1994).

46. Mann, J. Natural products as immunosuppressive agents. *Nat Prod Rep* **18**, 417–430 (2001).

47. Schwecke, T. *et al.* The biosynthetic gene cluster for the polyketide immunosuppressant rapamycin. *Proc National Acad Sci* **92**, 7839–7843 (1995).

48. Meggers, E. From conventional to unusual enzyme inhibitor scaffolds: the quest for target specificity. *Angewandte Chemie Int Ed Engl* **50**, 2442–8 (2011).

49. Fujita, E. & Nagao, Y. Tumor inhibitors having potential for interaction with mercapto enzymes and/or coenzymes. *Bioorg Chem* **6**, 287–309 (1977).

50. Cutler, H. G., Hill, R. A., Ward, B. G., Rohitha, B. H. & Stewart, A. Antimicrobial, Insecticidal, and Medicinal Properties of Natural Product Flavors and Fragrances. 51–66 (1996) doi:10.1021/bk-1996-0637.ch005.

51. Liu, H. & Begley, T. Natural product biosynthesis--a Renaissance. *Curr Opin Chem Biol* **17**, 529–31 (2013).

52. Bentley, S. D. *et al.* Complete genome sequence of the model actinomycete Streptomyces coelicolor A3(2). *Nature* **417**, 141–147 (2002).

53. Udwary, D. W. *et al.* Genome sequencing reveals complex secondary metabolome in the marine actinomycete Salinispora tropica. *Proc National Acad Sci* **104**, 10376–10381 (2007).

54. Davies, J. Small molecules: The lexicon of biodiversity. *J Biotechnol* **129**, 3–5 (2007).

55. Stewart, E. Growing Unculturable Bacteria. *J Bacteriol* **194**, 4151–4160 (2012).

56. Blunt, J. W., Copp, B. R., Keyzers, R. A., Munro, M. H. & Prinsep, M. R. Marine natural products. *Nat Prod Rep* **33**, 382–431 (2016).

57. Carroll, A. R., Copp, B. R., Davis, R. A., Keyzers, R. A. & Prinsep, M. R. Marine natural products. *Nat Prod Rep* **36**, 122–173 (2019).

58. Rego, A. *et al.* Diversity of Bacterial Biosynthetic Genes in Maritime Antarctica. *Microorg* **8**, 279 (2020).

59. Nichols, D. *et al.* Use of Ichip for High-Throughput In Situ Cultivation of "Uncultivable" Microbial Species. *Appl Environ Microb* **76**, 2445–2450 (2010).

60. Ling, L. L. *et al.* A new antibiotic kills pathogens without detectable resistance. *Nature* **517**, 455–459 (2015).

61. Culp, E. J. *et al.* Evolution-guided discovery of antibiotics that inhibit peptidoglycan remodelling. *Nature* 1–6 (2020) doi:10.1038/s41586-020-1990-9.

62. Stokes, J. M. *et al.* A Deep Learning Approach to Antibiotic Discovery. *Cell* **180**, 688-702.e13 (2020).

63. Donk, W. A. van der. Introduction: Unusual Enzymology in Natural Product Synthesis. *Chem Rev* **117**, 5223–5225 (2017).

64. Cimermancic, P. *et al.* Insights into secondary metabolism from a global analysis of prokaryotic biosynthetic gene clusters. *Cell* **158**, 412–21 (2014).

65. Gabius, H.-J. The Sugar Code: Why glycans are so important. *Bio Syst* **164**, 102–111 (2017).

66. Oldfield, E. & Lin, F.-Y. Terpene biosynthesis: modularity rules. *Angewandte Chemie Int Ed Engl* **51**, 1124–37 (2011).

67. Liang, M.-H., Zhu, J. & Jiang, J.-G. Carotenoids biosynthesis and cleavage related genes from bacteria to plants. *Crit Rev Food Sci* **58**, 0 (2017).

68. Srinivasan, K. & Buys, E. M. Insights into the role of bacteria in vitamin A biosynthesis: Future research opportunities. *Crit Rev Food Sci* 1–16 (2019) doi:10.1080/10408398.2018.1546670.

69. Klayman, D. Qinghaosu (artemisinin): an antimalarial drug from China. *Science* **228**, 1049–1055 (1985).

70. Booth, J. K. & Bohlmann, J. Terpenes in Cannabis sativa – From plant genome to humans. *Plant Sci* **284**, 67–72 (2019).

71. Hudson, G. A. & Mitchell, D. A. RiPP antibiotics: biosynthesis and engineering potential. *Curr Opin Microbiol* **45**, 61–69 (2018).

72. Arnison, P. G. *et al.* Ribosomally synthesized and post-translationally modified peptide natural products: overview and recommendations for a universal nomenclature. *Nat Prod Rep* **30**, 108–160 (2013).

73. Keller, U., Lang, M., Crnovcic, I., Pfennig, F. & Schauwecker, F. The actinomycin biosynthetic gene cluster of Streptomyces chrysomallus: a genetic hall of mirrors for synthesis of a molecule with mirror symmetry. *J Bacteriol* **192**, 2583–95 (2010).

74. Recktenwald, J. *et al.* Nonribosomal biosynthesis of vancomycin-type antibiotics: a heptapeptide backbone and eight peptide synthetase modules. *Microbiology+* **148**, 1105–1118 (2002).

75. Ikeda, H. & Ōmura, S. Avermectin Biosynthesis. *Chem Rev* **97**, 2591–2610 (1997).

76. Cane, D. E. Programming of erythromycin biosynthesis by a modular polyketide synthase. *J Biological Chem* **285**, 27517–23 (2010).

77. Park, S., Yoo, Y., Ban, Y.-H. & Yoon, Y. Biosynthesis of rapamycin and its regulation: past achievements and recent progress. *J Antibiotics* **63**, 434–441 (2010).

78. Julien, B. *et al.* Isolation and characterization of the epothilone biosynthetic gene cluster from Sorangium cellulosum. *Gene* **249**, 153–160 (2000).

79. Bayly, C. L. & Yadav, V. G. Towards Precision Engineering of Canonical Polyketide Synthase Domains: Recent Advances and Future Prospects. *Mol Basel Switz* **22**, 235 (2017).

80. Weissman, K. J. & Leadlay, P. F. Combinatorial biosynthesis of reduced polyketides. *Nat Rev Microbiol* **3**, 925–936 (2005).

81. Keasling, J. D. Synthetic biology for synthetic chemistry. *Acs Chem Biol* **3**, 64–76 (2008).

82. Ladner, C. C. & Williams, G. J. Harnessing natural product assembly lines: structure, promiscuity, and engineering. *J Ind Microbiol Biot* **43**, 371–87 (2015).

83. Burkhart, B. J., Kakkar, N., Hudson, G. A., Donk, W. A. van der & Mitchell, D. A. Chimeric Leader Peptides for the Generation of Non-Natural Hybrid RiPP Products. *Acs Central Sci* **3**, 629–638 (2017).

84. Smith, S. & Tsai, S.-C. The type I fatty acid and polyketide synthases : a tale of two megasynthases. *Natural Product Reports* **24**, 1041–1072 (2007).

85. Yadav, G., Gokhale, R. S. & Mohanty, D. Towards prediction of metabolic products of polyketide synthases: an in silico analysis. *Plos Comput Biol* **5**, e1000351 (2009).

86. Helfrich, E. J. N. *et al.* Automated structure prediction of trans-acyltransferase polyketide synthase products. *Nat Chem Biol* **15**, 813–821 (2019).

87. Sieber, S. A. & Marahiel, M. A. Molecular Mechanisms Underlying Nonribosomal Peptide Synthesis: Approaches to New Antibiotics. *Chem Rev* **105**, 715–738 (2005).

88. Meier, J. L. & Burkart, M. D. The chemical biology of modular biosynthetic enzymes. *Chem Soc Rev* **38**, 2012–45 (2009).

89. Dunn, B. J., Cane, D. E. & Khosla, C. Mechanism and Specificity of an Acyltransferase Domain from a Modular Polyketide Synthase. *Biochemistry* **52**, 1839–1841 (2013).

90. Katz, L. Methods in Enzymology. *Section I. Polyketides* **459**, 113–142 (2009).

91. Khosla, C., Tang, Y., Chen, A. Y., Schnarr, N. A. & Cane, D. E. Structure and Mechanism of the 6-Deoxyerythronolide B Synthase. *Annual Review of Biochemistry* **76**, 195–221 (2007).

92. Leadlay, P. F. Combinatorial approaches to polyketide biosynthesis. *Curr Opin Chem Biol* **1**, 162–168 (1997).

93. Crawford, J. M. & Townsend, C. A. New insights into the formation of fungal aromatic polyketides. *Nat Rev Microbiol* **8**, 879–889 (2010).

94. Chen, H. & Du, L. Iterative polyketide biosynthesis by modular polyketide synthases in bacteria. *Appl Microbiol Biot* **100**, 541–557 (2015).

95. Lim, Y., Go, M. & Yew, W. Exploiting the Biosynthetic Potential of Type III Polyketide Synthases. *Molecules* **21**, 806 (2016).

96. Hertweck, C., Luzhetskyy, A., Rebets, Y. & Bechthold, A. Type II polyketide synthases: gaining a deeper insight into enzymatic teamwork. *Nat Prod Rep* **24**, 162–190 (2007).

97. Zhang, W. & Tang, Y. Methods in Enzymology. *Section I. Polyketides* **459**, 367–393 (2009).

98. Burson, K. K. & Khosla, C. Dissecting the Chain Length Specificity in Bacterial Aromatic Polyketide Synthases using Chimeric Genes. *Tetrahedron* **56**, 9401–9408 (2000).

99. Keatinge-Clay, A. T., Maltby, D. A., Medzihradszky, K. F., Khosla, C. & Stroud, R. M. An antibiotic factory caught in action. *Nat Struct Mol Biol* **11**, 888–893 (2004).

100. Zhang, M. *et al.* Biosynthesis of trioxacarcin revealing a different starter unit and complex tailoring steps for type II polyketide synthase. *Chem Sci* **6**, 3440–3447 (2015).

101. Zhang, Z., Pan, H.-X. & Tang, G.-L. New insights into bacterial type II polyketide biosynthesis. *F1000research* **6**, 172 (2017).

102. Li, Y. & Müller, R. Non-modular polyketide synthases in myxobacteria. *Phytochemistry* **70**, 1850–1857 (2009).

103. Yu, D., Xu, F., Zeng, J. & Zhan, J. Type III polyketide synthases in natural product biosynthesis. *Iubmb Life* **64**, 285–95 (2012).

104. Parvez, A. *et al.* Novel Type III Polyketide Synthases Biosynthesize Methylated Polyketides in Mycobacterium marinum. *Sci Rep-uk* **8**, 6529 (2018).

105. Miyanaga, A., Funa, N., Awakawa, T. & Horinouchi, S. Direct transfer of starter substrates from type I fatty acid synthase to type III polyketide synthases in phenolic lipid synthesis. *P Natl Acad Sci Usa* **105**, 871–6 (2008).

106. Singh, M., Chaudhary, S. & Sareen, D. Non-ribosomal peptide synthetases: Identifying the cryptic gene clusters and decoding the natural product. *J Biosciences* **42**, 175–187 (2017).

107. Cortes, J., Haydock, S. F., Roberts, G. A., Bevitt, D. J. & Leadlay, P. F. An unusually large multifunctional polypeptide in the erythromycin-producing polyketide synthase of Saccharopolyspora erythraea. *Nature* **348**, 176–178 (1990).

108. Pfeifer, B. A., Admiraal, S. J., Gramajo, H., Cane, D. E. & Khosla, C. Biosynthesis of Complex Polyketides in a Metabolically Engineered Strain of E. coli. *Science* **291**, 1790–1792 (2001).

109. Donadio, S., Staver, M., McAlpine, J., Swanson, S. & Katz, L. Modular organization of genes required for complex polyketide biosynthesis. *Science* **252**, 675–679 (1991).

110. Donadio, S. & Katz, L. Organization of the enzymatic domains in the multifunctional polyketide synthase involved in erythromycin formation in Saccharopolyspora erythraea. *Gene* **111**, 51–60 (1992).

111. Cane, D. E. & Yang, C. C. Macrolide biosynthesis. 4. Intact incorporation of a chain-elongation intermediate into erythromycin. *J Am Chem Soc* **109**, 1255–1257 (1987).

112. Cane, D. E., Liang, T. C., Taylor, P. B., Chang, C. & Yang, C. C. Macrolide biosynthesis. 3. Stereochemistry of the chain-elongation steps of erythromycin biosynthesis. *J Am Chem Soc* **108**, 4957–4964 (1986).

113. Tang, Y., Kim, C.-Y., Mathews, I. I., Cane, D. E. & Khosla, C. The 2.7-Å crystal structure of a 194-kDa homodimeric fragment of the 6-deoxyerythronolide B synthase. *Proceedings of the National Academy of Sciences* **103**, 11124–11129 (2006).

114. Tang, Y., Chen, A. Y., Kim, C.-Y., Cane, D. E. & Khosla, C. Structural and Mechanistic Analysis of Protein Interactions in Module 3 of the 6-Deoxyerythronolide B Synthase. *Chem Biol* **14**, 931–943 (2007).

115. Weissman, K. J. The Structural Basis for Docking in Modular Polyketide Biosynthesis. *Chembiochem* **7**, 485–494 (2006).

116. Buchholz, T. J. *et al.* Structural Basis for Binding Specificity between Subclasses of Modular Polyketide Synthase Docking Domains. *Acs Chem Biol* **4**, 41–52 (2009).

117. Zhang, W., Zhou, L., Li, C., Deng, Z. & Qu, X. Rational engineering acyltransferase domain of modular polyketide synthase for expanding substrate specificity. *Methods Enzymol* **622**, 271–292 (2019).

118. Sherman, D. H. The Lego-ization of polyketide biosynthesis. *Nat Biotechnol* **23**, 1083–1084 (2005).

119. González-Lergier, J., Broadbelt, L. J. & Hatzimanikatis, V. Theoretical Considerations and Computational Analysis of the Complexity in Polyketide Synthesis Pathways. *J Am Chem Soc* **127**, 9930–9938 (2005).

120. McDaniel, R. *et al.* Multiple genetic modifications of the erythromycin polyketide synthase to produce a library of novel “unnatural” natural products. *Proc National Acad Sci* **96**, 1846–1851 (1999).

121. Menzella, H. G. *et al.* Combinatorial polyketide biosynthesis by de novo design and rearrangement of modular polyketide synthase genes. *Nat Biotechnol* **23**, 1171–1176 (2005).

122. DUTTON, C. J. *et al.* Novel avermectins produced by mutational biosynthesis. *J Antibiotics* **44**, 357–365 (1991).

123. Haydock, S. F. *et al.* Divergent sequence motifs correlated with the substrate specificity of (methyl)malonyl-CoA:acyl carrier protein transacylase domains in modular polyketide synthases. *Febs Lett* **374**, 246–248 (1995).

124. Kumar, P., Koppisch, A. T., Cane, D. E. & Khosla, C. Enhancing the Modularity of the Modular Polyketide Synthases: Transacylation in Modular Polyketide Synthases Catalyzed by Malonyl-CoA:ACP Transacylase. *J Am Chem Soc* **125**, 14307–14312 (2003).

125. Dunn, B. J., Watts, K. R., Robbins, T., Cane, D. E. & Khosla, C. Comparative Analysis of the Substrate Specificity of trans- versus cis-Acyltransferases of Assembly Line Polyketide Synthases. *Biochemistry-us* **53**, 3796–3806 (2014).

126. Walker, M. C. *et al.* Expanding the Fluorine Chemistry of Living Systems Using Engineered Polyketide Synthase Pathways. *Science* **341**, 1089–1094 (2013).

127. Cai, W. & Zhang, W. Engineering modular polyketide synthases for production of biofuels and industrial chemicals. *Curr Opin Biotech* **50**, 32–38 (2017).

128. Yuzawa, S., Eng, C. H., Katz, L. & Keasling, J. D. Broad Substrate Specificity of the Loading Didomain of the Lipomycin Polyketide Synthase. *Biochemistry-us* **52**, 3791–3793 (2013).

129. Broadhurst, W. R., Nietlispach, D., Wheatcroft, M. P., Leadlay, P. F. & Weissman, K. J. The Structure of Docking Domains in Modular Polyketide Synthases. *Chem Biol* **10**, 723–731 (2003).

130. Thomas, I., Martin, C. J., Wilkinson, C. J., Staunton, J. & Leadlay, P. F. Skipping in a Hybrid Polyketide Synthase. *Chem Biol* **9**, 781–787 (2002).

131. Moss, S. J., Martin, C. J. & Wilkinson, B. Loss of co-linearity by modular polyketide synthases: a mechanism for the evolution of chemical diversity. *Nat Prod Rep* **21**, 575 (2004).

132. Khosla, C., Kapur, S. & Cane, D. E. Revisiting the modularity of modular polyketide synthases. *Curr Opin Chem Biol* **13**, 135–143 (2009).

133. Aparicio, J., Caffrey, P., Marsden, A., Staunton, J. & Leadlay, P. Limited proteolysis and active-site studies of the first multienzyme component of the erythromycin-producing polyketide synthase. *J Biological Chem* **269**, 8524–8 (1994).

134. Tsai, C. S. *et al.* Crystal structure of the macrocycle-forming thioesterase domain of the erythromycin polyketide synthase: Versatility from a unique substrate channel. *Proc National Acad Sci* **98**, 14808–14813 (2001).

135. Keatinge-Clay, A. T. & Stroud, R. M. The Structure of a Ketoreductase Determines the Organization of the β-Carbon Processing Enzymes of Modular Polyketide Synthases. *Structure* **14**, 737–748 (2006).

136. Maier, T., Jenni, S. & Ban, N. Architecture of Mammalian Fatty Acid Synthase at 4.5 A Resolution. *Science* **311**, 1258–1262 (2006).

137. Alekseyev, V. Y., Liu, C. W., Cane, D. E., Puglisi, J. D. & Khosla, C. Solution structure and proposed domain-domain recognition interface of an acyl carrier protein domain from a modular polyketide synthase. *Protein Sci* **16**, 2093–2107 (2007).

138. Miyanaga, A., Iwasawa, S., Shinohara, Y., Kudo, F. & Eguchi, T. Structure-based analysis of the molecular interactions between acyltransferase and acyl carrier protein in vicenistatin biosynthesis. *P Natl Acad Sci Usa* **113**, 1802–7 (2016).

139. Keatinge-Clay, A. Crystal Structure of the Erythromycin Polyketide Synthase Dehydratase. *J Mol Biol* **384**, 941–953 (2008).

140. Zheng, J., Gay, D. C., Demeler, B., White, M. A. & Keatinge-Clay, A. T. Divergence of multimodular polyketide synthases revealed by a didomain structure. *Nat Chem Biol* **8**, 615 (2012).

141. Dutta, S. *et al.* Structure of a modular polyketide synthase. *Nature* **510**, 512–7 (2014).

142. Whicher, J. R. *et al.* Structural rearrangements of a polyketide synthase module during its catalytic cycle. *Nature* **510**, 560 (2014).

143. Edwards, A. L., Matsui, T., Weiss, T. M. & Khosla, C. Architectures of Whole-Module and Bimodular Proteins from the 6-Deoxyerythronolide B Synthase. *J Mol Biol* **426**, 2229–2245 (2014).

144. Davison, J. *et al.* Insights into the function of trans-acyl transferase polyketide synthases from the SAXS structure of a complete module. *Chem Sci* **5**, 3081–3095 (2014).

145. Dorival, J. *et al.* Characterization of Intersubunit Communication in the Virginiamycin trans -Acyl Transferase Polyketide Synthase. *J Am Chem Soc* **138**, 4155–4167 (2016).

146. Lowry, B., Li, X., Robbins, T., Cane, D. E. & Khosla, C. A Turnstile Mechanism for the Controlled Growth of Biosynthetic Intermediates on Assembly Line Polyketide Synthases. *Acs Central Sci* **2**, 14–20 (2016).

147. Robbins, T., Liu, Y.-C., Cane, D. E. & Khosla, C. Structure and mechanism of assembly line polyketide synthases. *Curr Opin Struc Biol* **41**, 10–18 (2016).

148. Weissman, K. J. Uncovering the structures of modular polyketide synthases. *Nat Prod Rep* **32**, 436–453 (2015).

149. Weissman, K. J. The structural biology of biosynthetic megaenzymes. *Nat Chem Biol* **11**, 660–670 (2015).

150. Klaus, M. & Grininger, M. Engineering strategies for rational polyketide synthase design. *Nat Prod Rep* **35**, 1070–1081 (2018).

151. Kalkreuter, E. & Williams, G. J. Engineering enzymatic assembly lines for the production of new antimicrobials. *Current opinion in microbiology* **45**, 140–148 (2018).

152. Maier, T., Leibundgut, M. & Ban, N. The Crystal Structure of a Mammalian Fatty Acid Synthase. *Science* **321**, 1315–1322 (2008).

153. Valenzano, C. R. *et al.* Stereospecificity of the dehydratase domain of the erythromycin polyketide synthase. *J Am Chem Soc* **132**, 14697–9 (2010).

154. Zhang, L. *et al.* Stereospecificity of Enoylreductase Domains from Modular Polyketide Synthases. *Acs Chem Biol* **13**, 871–875 (2018).

155. Koch, A. A. *et al.* A single active site mutation in the pikromycin thioesterase generates a more effective macrocyclization catalyst. *Journal of the American Chemical Society* (2017) doi:10.1021/jacs.7b06436.

156. Murphy, A. C., Hong, H., Vance, S., Broadhurst, R. W. & Leadlay, P. F. Broadening substrate specificity of a chain-extending ketosynthase through a single active-site mutation. *Chem Commun* **52**, 8373–8376 (2016).

157. Wang, F. *et al.* Structural and Functional Analysis of the Loading Acyltransferase from Avermectin Modular Polyketide Synthase. *Acs Chem Biol* **10**, 1017–1025 (2015).

158. Zheng, Q. *et al.* Enzyme-Dependent [4 + 2] Cycloaddition Depends on Lid-like Interaction of the N-Terminal Sequence with the Catalytic Core in PyrI4. *Cell Chem Biol* **23**, 352–360 (2016).

159. Kushnir, S. *et al.* Minimally Invasive Mutagenesis Gives Rise to a Biosynthetic Polyketide Library. *Angewandte Chemie Int Ed* **51**, 10664–10669 (2012).

160. Weissman, K. J. Genetic engineering of modular PKSs: from combinatorial biosynthesis to synthetic biology. *Nat Prod Rep* **33**, 203–230 (2016).

161. Yuzawa, S. *et al.* Heterologous Gene Expression of N-terminally truncated variants of LipPks1 Suggests a Functionally Critical Structural Motif in the N-terminus of Modular Polyketide Synthase. *ACS Chemical Biology* (2017) doi:10.1021/acschembio.7b00714.

162. Kalkreuter, E., CroweTipton, J. M., Lowell, A. N., Sherman, D. H. & Williams, G. J. Engineering the Substrate Specificity of a Modular Polyketide Synthase for Installation of Consecutive Non-Natural Extender Units. *J Am Chem Soc* **141**, 1961–1969 (2019).

163. Hagen, A. *et al.* Engineering a Polyketide Synthase for In Vitro Production of Adipic Acid. *Acs Synth Biol* **5**, 21–7 (2015).

164. Niquille, D. L. *et al.* Nonribosomal biosynthesis of backbone-modified peptides. *Nat Chem* **10**, 282–287 (2017).

165. Owen, J. G., Calcott, M. J., Robins, K. J. & Ackerley, D. F. Generating Functional Recombinant NRPS Enzymes in the Laboratory Setting via Peptidyl Carrier Protein Engineering. *Cell Chem Biol* **23**, 1395–1406 (2016).

166. Kasey, C. M., Zerrad, M., Li, Y., Cropp, T. A. & Williams, G. J. Development of Transcription Factor-Based Designer Macrolide Biosensors for Metabolic Engineering and Synthetic Biology. *Acs Synth Biol* **7**, 227–239 (2017).

167. Wlodek, A. *et al.* Diversity oriented biosynthesis via accelerated evolution of modular gene clusters. *Nat Commun* **8**, 1206 (2017).

168. Zhang, L. *et al.* Characterization of Giant Modular PKSs Provides Insight into Genetic Mechanism for Structural Diversification of Aminopolyol Polyketides. *Angewandte Chemie Int Ed* **56**, 1740–1745 (2017).

169. Keatinge-Clay, A. T. Polyketide Synthase Modules Redefined. *Angewandte Chemie Int Ed Engl* **56**, 4658–4660 (2017).

170. Miyazawa, T., Hirsch, M., Zhang, Z. & Keatinge-Clay, A. T. An in vitro platform for engineering and harnessing modular polyketide synthases. *Nat Commun* **11**, 80 (2020).

171. Tibrewal, N. & Tang, Y. Biocatalysts for natural product biosynthesis. *Annu Rev Chem Biomol* **5**, 347–66 (2014).

172. Li, Z. *et al.* Oxidative biotransformations using oxygenases. *Curr Opin Chem Biol* **6**, 136–144 (2002).

173. Palfey, B. A. & McDonald, C. A. Control of catalysis in flavin-dependent monooxygenases. *Arch Biochem Biophys* **493**, 26–36 (2010).

174. Yeh, E., Garneau, S. & Walsh, C. T. Robust in vitro activity of RebF and RebH, a two-component reductase/halogenase, generating 7-chlorotryptophan during rebeccamycin biosynthesis. *Proc National Acad Sci* **102**, 3960–3965 (2005).

175. Glenn, W. S., Nims, E. & O'Connor, S. E. Reengineering a Tryptophan Halogenase To Preferentially Chlorinate a Direct Alkaloid Precursor. *J Am Chem Soc* **133**, 19346–19349 (2011).

176. Payne, J. T., Andorfer, M. C. & Lewis, J. C. Regioselective Arene Halogenation using the FAD-Dependent Halogenase RebH. *Angewandte Chemie Int Ed* **52**, 5271–5274 (2013).

177. Zhu, T., Cheng, X., Liu, Y., Deng, Z. & You, D. Deciphering and engineering of the final step halogenase for improved chlortetracycline biosynthesis in industrial Streptomyces aureofaciens. *Metab Eng* **19**, 69–78 (2013).

178. Benjdia, A., Balty, C. & Berteau, O. Radical SAM Enzymes in the Biosynthesis of Ribosomally Synthesized and Post-translationally Modified Peptides (RiPPs). *Front Chem* **5**, 87 (2017).

179. Pierre, S. *et al.* Thiostrepton tryptophan methyltransferase expands the chemistry of radical SAM enzymes. *Nat Chem Biol* **8**, 957–959 (2012).

180. Eustáquio, A. S., Pojer, F., Noel, J. P. & Moore, B. S. Discovery and characterization of a marine bacterial SAM-dependent chlorinase. *Nat Chem Biol* **4**, 69–74 (2007).

181. Eustaquio, A. S. & Moore, B. S. Mutasynthesis of Fluorosalinosporamide, a Potent and Reversible Inhibitor of the Proteasome. *Angewandte Chemie Int Ed* **47**, 3936–3938 (2008).

182. Xie, X. & Tang, Y. Efficient Synthesis of Simvastatin by Use of Whole-Cell Biocatalysis. *Appl Environ Microb* **73**, 2054–2060 (2007).

183. Gao, X. *et al.* Directed Evolution and Structural Characterization of a Simvastatin Synthase. *Chem Biol* **16**, 1064–1074 (2009).

184. Zhang, C., Albermann, C., Fu, X. & Thorson, J. S. The in Vitro Characterization of the Iterative Avermectin Glycosyltransferase AveBI Reveals Reaction Reversibility and Sugar Nucleotide Flexibility. *J Am Chem Soc* **128**, 16420–16421 (2006).

185. Williams, G. J., Goff, R. D., Zhang, C. & Thorson, J. S. Optimizing Glycosyltransferase Specificity via "Hot Spot" Saturation Mutagenesis Presents a Catalyst for Novobiocin Glycorandomization. *Chem Biol* **15**, 393–401 (2008).

186. Gantt, R. W., Peltier-Pain, P., Singh, S., Zhou, M. & Thorson, J. S. Broadening the scope of glycosyltransferase-catalyzed sugar nucleotide synthesis. *Proc National Acad Sci* **110**, 7648–7653 (2013).

187. Parajuli, P. *et al.* Enzymatic synthesis of epothilone A glycosides. *Amb Express* **4**, 31 (2014).

188. Dickschat, J. S. Bacterial terpene cyclases. *Nat Prod Rep* **33**, 87–110 (2016).

189. Patel, A. *et al.* Dynamically complex [6+ 4] and [4+ 2] cycloadditions in the biosynthesis of spinosyn A. *Journal of the American Chemical Society* **138**, 3631–3634 (2016).

190. Fage, C. D. *et al.* The structure of SpnF, a standalone enzyme that catalyzes [4 + 2] cycloaddition. *Nat Chem Biol* **11**, 256–8 (2015).

191. Sultana, A. *et al.* Structure of the polyketide cyclase SnoaL reveals a novel mechanism for enzymatic aldol condensation. *Embo J* **23**, 1911–1921 (2004).

192. Bister, B. *et al.* Abyssomicin C—A Polycyclic Antibiotic from a MarineVerrucosispora Strain as an Inhibitor of thep-Aminobenzoic Acid/Tetrahydrofolate Biosynthesis Pathway. *Angewandte Chemie Int Ed* **43**, 2574–2576 (2004).

193. Nouioui, I. *et al.* Genome-Based Taxonomic Classification of the Phylum Actinobacteria. *Front Microbiol* **9**, 2007 (2018).

194. Riedlinger, J. *et al.* Abyssomicins, inhibitors of the para-aminobenzoic acid pathway produced by the marine Verrucosispora strain AB-18-032. *The Journal of antibiotics* **57**, 271–9 (2004).

195. Nicolaou, K. C. & Harrison, S. T. Total Synthesis of Abyssomicin C and atrop-Abyssomicin C. *Angew Chem-ger Edit* **118**, 3334–3338 (2006).

196. Keller, S., Schadt, H. S., Ortel, I. & Süssmuth, R. D. Action of atrop-Abyssomicin C as an Inhibitor of 4-Amino-4-deoxychorismate Synthase PabB. *Angewandte Chemie Int Ed* **46**, 8284–8286 (2007).

197. Keller, S. *et al.* Abyssomicins G and H and atrop-Abyssomicin C from the Marine Verrucosispora Strain AB-18-032†. *J Antibiotics* **60**, 391–394 (2007).

198. Vieweg, L., Reichau, S., Schobert, R., Leadlay, P. F. & Süssmuth, R. D. Recent advances in the field of bioactive tetronates. *Nat Prod Rep* **31**, 1554–84 (2014).

199. Bihelovic, F., Karadzic, I., Matovic, R. & Saicic, R. N. Total synthesis and biological evaluation of (−)-atrop–abyssomicin C. *Organic & Biomolecular Chemistry* **11**, 5413–5424 (2013).

200. Savic, V. Studies in Natural Products Chemistry. *Stud Nat Prod Chem* **40**, 133–172 (2013).

201. Sadaka, C. *et al.* Review on Abyssomicins: Inhibitors of the Chorismate Pathway and Folate Biosynthesis. *Molecules* **23**, 1371 (2018).

202. Dosselaere, F. & Vanderleyden, J. A Metabolic Node in Action: Chorismate-Utilizing Enzymes in Microorganisms. *Crit Rev Microbiol* **27**, 75–131 (2001).

203. Wright, P. M., Seiple, I. B. & Myers, A. G. The evolving role of chemical synthesis in antibacterial drug discovery. *Angewandte Chemie Int Ed Engl* **53**, 8840–69 (2014).

204. Sköld, O. Resistance to trimethoprim and sulfonamides. *Vet Res* **32**, 261–273 (2001).

205. Sköld, O. Sulfonamide resistance: mechanisms and trends. *Drug Resist Update* **3**, 155–160 (2000).

206. Huovinen, P. Trimethoprim resistance. *Antimicrob Agents Ch* **31**, 1451–1456 (1987).

207. Freundlich, J. S. *et al.* The Abyssomicin C family as in vitro inhibitors of Mycobacterium tuberculosis. *Tuberculosis* **90**, 298–300 (2010).

208. Wang, Q. *et al.* Abyssomicins from the South China Sea Deep-Sea Sediment Verrucosispora sp.: Natural Thioether Michael Addition Adducts as Antitubercular Prodrugs . *Angewandte Chemie International Edition* **52**, 1231–1234 (2013).

209. Song, Y. *et al.* Neoabyssomicins A–C, polycyclic macrolactones from the deep-sea derived Streptomyces koyangensis SCSIO 5802. *Tetrahedron* **73**, 5366–5372 (2017).

210. Tu, J. *et al.* Characterization and heterologous expression of the neoabyssomicin/abyssomicin biosynthetic gene cluster from Streptomyces koyangensis SCSIO 5802. *Microbial Cell Factories* **17**, 28 (2018).

211. León, B. *et al.* Abyssomicin 2 Reactivates Latent HIV-1 by a PKC- and HDAC-Independent Mechanism. *Org Lett* **17**, 262–265 (2015).

212. Richman, D. D. *et al.* The challenge of finding a cure for HIV infection. *Sci New York N Y* **323**, 1304–7 (2009).

213. Zhang, J. *et al.* A new abyssomicin polyketide with anti-influenza A virus activity from a marine-derived Verrucosispora sp. MS100137. *Appl Microbiol Biot* **104**, 1533–1543 (2020).

214. Matovic, R., Bihelovic, F., Gruden-Pavlovic, M. & Saicic, R. N. Total synthesis and biological evaluation of atrop-O-benzyl-desmethylabyssomicin C. *Org Biomol Chem* **12**, 7682–7685 (2014).

215. Monjas, L. *et al.* Synthesis and biological evaluation of truncated derivatives of abyssomicin C as antibacterial agents. *Beilstein J Org Chem* **15**, 1468–1474 (2019).

216. Zapf, C. W., Harrison, B. A., Drahl, C. & Sorensen, E. J. A Diels-Alder Macrocyclization Enables an Efficient Asymmetric Synthesis of the Antibacterial Natural Product Abyssomicin C. *Angewandte Chemie Int Ed* **44**, 6533–6537 (2005).

217. Nicolaou, K. C. & Harrison, S. T. Total Synthesis of Abyssomicin C, Atrop-abyssomicin C, and Abyssomicin D: Implications for Natural Origins of Atrop-abyssomicin C. *J Am Chem Soc* **129**, 429–440 (2007).

218. Couladouros, E. A., Bouzas, E. A. & Magos, A. D. Formal synthesis of Abyssomicin C. *Tetrahedron* **62**, 5272–5279 (2006).

219. Snider, B. B. & Zou, Y. Synthesis of the Carbocyclic Skeleton of Abyssomicins C and D. *Org Lett* **7**, 4939–4941 (2005).

220. Gottardi, E. M. *et al.* Abyssomicin Biosynthesis: Formation of an Unusual Polyketide, Antibiotic-Feeding Studies and Genetic Analysis. *ChemBioChem* **12**, 1401–1410 (2011).

221. Jia, X.-Y. *et al.* Genetic Characterization of the Chlorothricin Gene Cluster as a Model for Spirotetronate Antibiotic Biosynthesis. *Chem Biol* **13**, 575–585 (2006).

222. Zhang, H. *et al.* Elucidation of the Kijanimicin Gene Cluster: Insights into the Biosynthesis of Spirotetronate Antibiotics and Nitrosugars. *J Am Chem Soc* **129**, 14670–14683 (2007).

223. Maschio, L. *et al.* Cloning, expression, and purification of intact polyketide synthase modules. *Methods in Enzymology* 63–82 (2019) doi:10.1016/bs.mie.2018.12.018.

224. Byrne, M. J. *et al.* The Catalytic Mechanism of a Natural Diels-Alderase Revealed in Molecular Detail. *Journal of the American Chemical Society* **138**, 6095–8 (2016).

225. Lees, N. R. *et al.* An Esterase-like Lyase Catalyzes Acetate Elimination in Spirotetronate/Spirotetramate Biosynthesis. *Angew Chem-ger Edit* **131**, 2327–2331 (2019).

226. Li, Q. *et al.* AbmV Catalyzes Tandem Ether Installation and Hydroxylation during Neoabyssomicin/Abyssomicin Biosynthesis. *Org Lett* **20**, 4854–4857 (2018).

227. Ji, X. *et al.* A Luciferase-Like Monooxygenase and Flavin Reductase Pair AbmE2/AbmZ Catalyzes Baeyer–Villiger Oxidation in Neoabyssomicin Biosynthesis. *Acs Catal* **10**, 2591–2595 (2020).

228. Frommer, W. B. & Ninnemann, O. Heterologous Expression of Genes in Bacterial, Fungal, Animal, and Plant Cells. *Annu Rev Plant Phys* **46**, 419–444 (1995).

229. Meinke, D. W., Cherry, J. M., Dean, C., Rounsley, S. D. & Koornneef, M. Arabidopsis thaliana: A Model Plant for Genome Analysis. *Science* **282**, 662–682 (1998).

230. Wienkoop, S., Baginsky, S. & Weckwerth, W. Arabidopsis thaliana as a model organism for plant proteome research. *J Proteomics* **73**, 2239–2248 (2010).

231. Blümel, M., Dally, N. & Jung, C. Flowering time regulation in crops — what did we learn from Arabidopsis? *Curr Opin Biotech* **32**, 121–129 (2015).

232. Haughn, G. W., Smith, J., Mazur, B. & Somerville, C. Transformation with a mutant Arabidopsis acetolactate synthase gene renders tobacco resistant to sulfonylurea herbicides. *Mol Gen Genetics Mgg* **211**, 266–271 (1988).

233. Eenennaam, A. L. V. *et al.* Engineering Vitamin E Content: From Arabidopsis Mutant to Soy Oil. *Plant Cell* **15**, 3007–3019 (2003).

234. Bedford, N. L. & Hoekstra, H. E. Peromyscus mice as a model for studying natural variation. *Elife* **4**, e06813 (2015).

235. Buselmaier, W., Vierling, Th., Balzereit, W. & Schwegler, H. Genetic analysis of avoidance learning by means of different psychological testing systems with inbred mice as model organisms. *Psychological Res* **43**, 317–333 (1981).

236. Maschio, A. *et al.* Transgenic mice carrying the guinea-pig α-lactalbumin gene transcribe milk protein genes in their sebaceous glands during lactation. *Biochem J* **275**, 459–467 (1991).

237. Miltenburg, M. H. van & Jonkers, J. Using genetically engineered mouse models to validate candidate cancer genes and test new therapeutic approaches. *Curr Opin Genet Dev* **22**, 21–7 (2012).

238. Sharma, S. V., Haber, D. A. & Settleman, J. Cell line-based platforms to evaluate the therapeutic efficacy of candidate anticancer agents. *Nat Rev Cancer* **10**, 241–53 (2010).

239. Wolf, D. Cellular control in the eukaryotic cell through action of proteinases: the yeast Saccharomyces cerevisiae as a model organism. *Microbiological sciences* **3**, 107—11, 114 (1986).

240. Blount, Z. D. The unexhausted potential of E. coli. *Elife* **4**, e05826 (2015).

241. Ankeny, R. A. & Leonelli, S. What's so special about model organisms? *Stud Hist Philosophy Sci Part* **42**, 313–323 (2011).

242. Yang, S.-H. *et al.* Towards a transgenic model of Huntington's disease in a non-human primate. *Nature* **453**, 921–4 (2008).

243. Parikh, A., Gillam, E. M. J. & Guengerich, F. P. Drug metabolism by Escherichia coli expressing human cytochromes P450. *Nat Biotechnol* **15**, 784–788 (1997).

244. Brian, W. R. *et al.* Catalytic activities of human liver cytochrome P-450 IIIA4 expressed in Saccharomyces cerevisiae. *Biochemistry-us* **29**, 11280–11292 (1990).

245. McKENZIE, E. *et al.* Biochemical characterization of the active heterodimer form of human heparanase (Hpa1) protein expressed in insect cells. *Biochem J* **373**, 423–435 (2003).

246. Tzeng, E., Billiar, T. R., Robbins, P. D., Loftus, M. & Stuehr, D. J. Expression of human inducible nitric oxide synthase in a tetrahydrobiopterin (H4B)-deficient cell line: H4B promotes assembly of enzyme subunits into an active dimer. *Proc National Acad Sci* **92**, 11771–11775 (1995).

247. Fridman, R. *et al.* Expression of human recombinant 72 kDa gelatinase and tissue inhibitor of metalloproteinase-2 (TIMP-2): characterization of complex and free enzyme. *Biochem J* **289**, 411–416 (1993).

248. Wenzel, S. C. & Müller, R. Recent developments towards the heterologous expression of complex bacterial natural product biosynthetic pathways. *Current Opinion in Biotechnology* **16**, 594–606 (2005).

249. Wu, J. *et al.* Toward improvement of erythromycin A production in an industrial Saccharopolyspora erythraea strain via facilitation of genetic manipulation with an artificial attB site for specific recombination. *Appl Environ Microb* **77**, 7508–16 (2011).

250. Minas, W., Brunker, P., Kallio, P. T. & Bailey, J. E. Improved Erythromycin Production in a Genetically Engineered Industrial Strain of Saccharopolyspora erythraea. *Biotechnol Progr* **14**, 561–566 (1998).

251. Jiang, M. & Pfeifer, B. A. Metabolic and pathway engineering to influence native and altered erythromycin production through E. coli. *Metab Eng* **19**, 42–49 (2013).

252. Peiru, S., Menzella, H. G., Rodriguez, E., Carney, J. & Gramajo, H. Production of the Potent Antibacterial Polyketide Erythromycin C in Escherichia coli. *Appl Environ Microb* **71**, 2539–2547 (2005).

253. Fortman, J. L. & Sherman, D. H. Utilizing the Power of Microbial Genetics to Bridge the Gap Between the Promise and the Application of Marine Natural Products. *Chembiochem* **6**, 960–978 (2005).

254. Rutledge, P. J. & Challis, G. L. Discovery of microbial natural products by activation of silent biosynthetic gene clusters. *Nat Rev Microbiol* **13**, 509–523 (2015).

255. Komatsu, M., Uchiyama, T., Omura, S., Cane, D. E. & Ikeda, H. Genome-minimized Streptomyces host for the heterologous expression of secondary metabolism. *Proc National Acad Sci* **107**, 2646–2651 (2010).

256. Komatsu, M. *et al.* Engineered Streptomyces avermitilis Host for Heterologous Expression of Biosynthetic Gene Cluster for Secondary Metabolites. *Acs Synth Biol* **2**, 384–396 (2013).

257. Liu, Q. *et al.* Development of Streptomyces sp. FR-008 as an emerging chassis. *Synthetic Syst Biotechnology* **1**, 207–214 (2016).

258. McDaniel, R., Ebert-Khosla, S., Hopwood, D. & Khosla, C. Engineered biosynthesis of novel polyketides. *Science* **262**, 1546–1550 (1993).

259. Ziermann, R. & Betlach, M. C. Recombinant Polyketide Synthesis in Streptomyces : Engineering of Improved Host Strains. *Biotechniques* **26**, 106–110 (1999).

260. Kieser, T., Bibb, M. J., Buttner, M. J., Chater, K. F. & Hopwood, D. A. *Practical streptomyces genetics*. vol. 291 (John Innes Foundation Norwich, 2000).

261. Zaburannyi, N., Rabyk, M., Ostash, B., Fedorenko, V. & Luzhetskyy, A. Insights into naturally minimised Streptomyces albus J1074 genome. *Bmc Genomics* **15**, 97 (2014).

262. Tyo, K. E. J., Kocharin, K. & Nielsen, J. Toward design-based engineering of industrial microbes. *Curr Opin Microbiol* **13**, 255–62 (2010).

263. Zhang, M. M., Wang, Y., Ang, E. L. & Zhao, H. Engineering microbial hosts for production of bacterial natural products. *Nat Prod Rep* **33**, 963–987 (2016).

264. Yuet, K. P. & Khosla, C. Challenges and opportunities for engineering assembly-line polyketide biosynthesis in Escherichia coli. *Metabolic Eng Commun* e00106 (2019) doi:10.1016/j.mec.2019.e00106.

265. Lau, J., Tran, C., Licari, P. & Galazzo, J. Development of a high cell-density fed-batch bioprocess for the heterologous production of 6-deoxyerythronolide B in Escherichia coli. *J Biotechnol* **110**, 95–103 (2004).

266. Zhang, G., Li, Y., Fang, L. & Pfeifer, B. A. Tailoring pathway modularity in the biosynthesis of erythromycin analogs heterologously engineered in *E. coli*. *Science Advances* **1**, e1500077 (2015).

267. Chen, A. Y., Schnarr, N. A., Kim, C.-Y., Cane, D. E. & Khosla, C. Extender Unit and Acyl Carrier Protein Specificity of Ketosynthase Domains of the 6-Deoxyerythronolide B Synthase. *J Am Chem Soc* **128**, 3067–3074 (2006).

268. Watanabe, K., Wang, C. C. C., Boddy, C. N., Cane, D. E. & Khosla, C. Understanding Substrate Specificity of Polyketide Synthase Modules by Generating Hybrid Multimodular Synthases. *J Biol Chem* **278**, 42020–42026 (2003).

269. Ma, S. M. & Tang, Y. Biochemical characterization of the minimal polyketide synthase domains in the lovastatin nonaketide synthase LovB: Minimal polyketide synthase domains in LovB. *Febs J* **274**, 2854–2864 (2007).

270. Li, Y., Xu, W. & Tang, Y. Classification, prediction, and verification of the regioselectivity of fungal polyketide synthase product template domains. *J Biological Chem* **285**, 22764–73 (2010).

271. Parascandolo, J. S. *et al.* Insights into 6-Methylsalicylic Acid Bio-assembly by Using Chemical Probes. *Angewandte Chemie Int Ed Engl* **55**, 3463–7 (2016).

272. Xu, W., Cai, X., Jung, M. E. & Tang, Y. Analysis of intact and dissected fungal polyketide synthase-nonribosomal peptide synthetase in vitro and in Saccharomyces cerevisiae. *J Am Chem Soc* **132**, 13604–7 (2010).

273. Gustafsson, C., Govindarajan, S. & Minshull, J. Codon bias and heterologous protein expression. *Trends Biotechnol* **22**, 346–353 (2004).

274. Lithwick, G. & Margalit, H. Hierarchy of Sequence-Dependent Features Associated With Prokaryotic Translation. *Genome Res* **13**, 2665–2673 (2003).

275. Bird, L. E. *et al.* Application of In-Fusion™ Cloning for the Parallel Construction of E. coli Expression Vectors. *Methods Mol Biology Clifton N J* **1116**, 209–234 (2013).

276. Bachmann, B. O. & Ravel, J. Methods for In Silico Prediction of Microbial Polyketide and Nonribosomal Peptide Biosynthetic Pathways from DNA Sequence Data. *Methods Enzymol* **458**, 181–217 (2009).

277. Hughes, R. A. & Ellington, A. D. Synthetic DNA Synthesis and Assembly: Putting the Synthetic in Synthetic Biology. *Csh Perspect Biol* **9**, a023812 (2017).

278. Berrow, N. S. *et al.* A versatile ligation-independent cloning method suitable for high-throughput expression screening applications. *Nucleic Acids Research* **35**, e45–e45 (2007).

279. Engler, C., Gruetzner, R., Kandzia, R. & Marillonnet, S. Golden Gate Shuffling: A One-Pot DNA Shuffling Method Based on Type IIs Restriction Enzymes. *Plos One* **4**, e5553 (2009).

280. Engler, C. & Marillonnet, S. Methods in Molecular Biology. *Methods Mol Biology Clifton N J* **1073**, 141–156 (2013).

281. Engler, C. & Marillonnet, S. Golden Gate cloning. *Methods Mol Biology Clifton N J* **1116**, 119–31 (2014).

282. EAGON, R. G. Pseudomonas natriegens, a marine bacterium with a generation time of less than 10 minutes. *J Bacteriol* **83**, 736–7 (1962).

283. Weinstock, M. T., Hesek, E. D., Wilson, C. M. & Gibson, D. G. Vibrio natriegens as a fast-growing host for molecular biology. *Nat Methods* **13**, 849–51 (2016).

284. Beck, B. J., Aldrich, C. C., Fecik, R. A., Reynolds, K. A. & Sherman, D. H. Iterative chain elongation by a pikromycin monomodular polyketide synthase. *Journal of the American Chemical Society* **125**, 4682–3 (2003).

285. Phillips, G. N. Comparison of the dynamics of myoglobin in different crystal forms. *Biophys J* **57**, 381–383 (1990).

286. Sugase, K., Dyson, H. J. & Wright, P. E. Mechanism of coupled folding and binding of an intrinsically disordered protein. *Nature* **447**, 1021–1025 (2007).

287. Nguyen, T. *et al.* Exploiting the mosaic structure of trans-acyltransferase polyketide synthases for natural product discovery and pathway dissection. *Nat Biotechnol* **26**, 225–233 (2008).

288. Ridley, C. P., Lee, H. & Khosla, C. Evolution of polyketide synthases in bacteria. *P Natl Acad Sci Usa* **105**, 4595–600 (2008).

289. Asturias, F. J. *et al.* Structure and molecular organization of mammalian fatty acid synthase. *Nat Struct Mol Biol* **12**, 225–232 (2005).

290. Long, F., Vagin, A. A., Young, P. & Murshudov, G. N. BALBES : a molecular-replacement pipeline. *Acta Crystallogr Sect D Biological Crystallogr* **64**, 125–132 (2007).

291. Winn, M. D. *et al.* Overview of the CCP 4 suite and current developments. *Acta Crystallogr Sect D Biological Crystallogr* **67**, 235–242 (2011).

292. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr Sect D Biological Crystallogr* **66**, 486–501 (2010).

293. Maia, F. R. N. C., Szöke, A., DeLano, W. & Spoel, D. van der. Interactive visualization of electron density slices. *J Appl Crystallogr* **38**, 563–565 (2005).

294. Scheres, S. H. W. RELION: implementation of a Bayesian approach to cryo-EM structure determination. *J Struct Biol* **180**, 519–30 (2012).

295. Gay, D. C. *et al.* A Close Look at a Ketosynthase from a Trans-Acyltransferase Modular Polyketide Synthase. *Structure* **22**, 444–451 (2014).

296. Olsen, J. G. *et al.* The X-ray crystal structure of β-ketoacyl [acyl carrier protein] synthase I. *Febs Lett* **460**, 46–52 (1999).

297. Gay, D. C. *et al.* The LINKS motif zippers trans-acyltransferase polyketide synthase assembly lines into a biosynthetic megacomplex. *J Struct Biol* **193**, 196–205 (2016).

298. Wettstein-Knowles, P. von, Olsen, J. G., McGuire, K. A. & Henriksen, A. Fatty acid synthesis. *Febs J* **273**, 695–710 (2006).

299. Palovcak, E. *et al.* A simple and robust procedure for preparing graphene-oxide cryo-EM grids. *J Struct Biol* **204**, 80–84 (2018).

300. Sgro, G. G. & Costa, T. R. D. Cryo-EM Grid Preparation of Membrane Protein Samples for Single Particle Analysis. *Frontiers Mol Biosci* **5**, 74 (2018).

301. Auch, A. F., Klenk, H.-P. & Göker, M. Standard operating procedure for calculating genome-to-genome distances based on high-scoring segment pairs. *Stand Genomic Sci* **2**, 142–8 (2010).

302. Iancu, C. V. *et al.* Electron cryotomography sample preparation using the Vitrobot. *Nat Protoc* **1**, 2813–2819 (2006).

303. Kapur, S., Chen, A. Y., Cane, D. E. & Khosla, C. Molecular recognition between ketosynthase and acyl carrier protein domains of the 6-deoxyerythronolide B synthase. *Proc National Acad Sci* **107**, 22066–22071 (2010).

304. Nicolaou, K. C., Snyder, S. A., Montagnon, T. & Vassilikogiannakis, G. The Diels-Alder Reaction in Total Synthesis. *Angewandte Chemie Int Ed* **41**, 1668–1698 (2002).

305. Diels, O. & Alder, K. Synthesen in der hydroaromatischen Reihe. *Justus Liebigs Annalen der Chemie* **460**, 98–122 (1928).

306. Medicine, U. S. N. L. of. PubMed Entries for "diels-alder." (2020).

307. Klas, K., Tsukamoto, S., Sherman, D. H. & Williams, R. M. Natural Diels-Alderases: Elusive and Irresistable. *The Journal of organic chemistry* **80**, 11672–85 (2015).

308. Townsend, C. A. A "Diels-Alderase" at Last. *Chembiochem* **12**, 2267–2269 (2011).

309. Oliveira, B. L., Guo, Z. & Bernardes, G. J. L. Inverse electron demand Diels-Alder reactions in chemical biology. *Chem Soc Rev* **46**, 4895–4950 (2017).

310. Bear, B. R., Sparks, S. M. & Shea, K. J. The Type 2 Intramolecular Diels-Alder Reaction: Synthesis and Chemistry of Bridgehead Alkenes. *Angewandte Chemie (International ed. in English)* **40**, 820–849 (2001).

311. Funel, J. & Abele, S. Industrial Applications of the Diels–Alder Reaction. *Angewandte Chemie Int Ed* **52**, 3822–3863 (2013).

312. García, J. I., Mayoral, J. A. & Salvatella, L. Do Secondary Orbital Interactions Really Exist? *Accounts Chem Res* **33**, 658–664 (2000).

313. Hoffmann, R. & Woodward, R. B. Orbital Symmetries and endo-exo Relationships in Concerted Cycloaddition Reactions. *J Am Chem Soc* **87**, 4388–4389 (1965).

314. Constantino, A. F., Francisco, C. S., Cubides-Roman, D. C. & Jr, V. L. Hetero-Diels-Alder Reactions in the Synthesis of Biologically Active Nitrogen Compounds: A Review. *Curr Org Synth* **15**, 84–104 (2018).

315. Gong, P.-X., Li, H.-J., Wang, M., Cheng, Y.-F. & Wu, Y.-C. Synthesis of natural product inulavosin via Ga(OTf)3-Catalyzed Hetero Diels-Alder Dimerization of salicyl alcohol derivative. *Nat Prod Res* **33**, 2911–2916 (2018).

316. Yang, B. & Gao, S. Recent advances in the application of Diels-Alder reactions involving o-quinodimethanes, aza-o-quinone methides and o-quinone methides in natural product total synthesis. *Chem Soc Rev* **47**, 7926–7953 (2018).

317. Li, X., Han, J., Jones, A. X. & Lei, X. Chiral Boron Complex-Promoted Asymmetric Diels-Alder Cycloaddition and Its Application in Natural Product Synthesis. *J Org Chem* **81**, 458–68 (2015).

318. Settle, A. E. *et al.* Heterogeneous Diels–Alder catalysis for biomass-derived aromatic compounds. *Green Chem* **19**, 3468–3492 (2017).

319. Aragonès, A. C. *et al.* Electrostatic catalysis of a Diels–Alder reaction. *Nature* **531**, 88–91 (2016).

320. Jeon, B., Wang, S.-A., Ruszczycky, M. W. & Liu, H. Natural [4 + 2]-Cyclases. *Chemical Reviews* (2016) doi:10.1021/acs.chemrev.6b00578.

321. McGrath, N. A., Brichacek, M. & Njardarson, J. T. A Graphical Journey of Innovative Organic Architectures That Have Improved Our Lives. *J Chem Educ* **87**, 1348–1349 (2010).

322. Narula, A. P. S. ACS Symposium Series. 1–16 (2019) doi:10.1021/bk-2019-1321.ch001.

323. Houk, K. N. & Strozier, R. W. On Lewis acid catalysis of Diels-Alder reactions. *Journal of the American Chemical Society* (1973).

324. Oikawa, H., Katayama, K., Suzuki, Y. & Ichihara, A. Enzymatic activity catalysing exo-selective Diels–Alder reaction in solanapyrone biosynthesis. *J Chem Soc Chem Commun* **0**, 1321–1322 (1995).

325. Auclair, K. *et al.* Lovastatin Nonaketide Synthase Catalyzes an Intramolecular Diels−Alder Reaction of a Substrate Analogue. *J Am Chem Soc* **122**, 11519–11520 (2000).

326. Romesberg, F. E., Spiller, B., Schultz, P. G. & Stevens, R. C. Immunological Origins of Binding and Catalysis in a Diels-Alderase Antibody. *Science* **279**, 1929–1933 (1998).

327. Stuhlmann, F. & Jäschke, A. Characterization of an RNA Active Site: Interactions between a Diels−Alderase Ribozyme and Its Substrates and Products. *J Am Chem Soc* **124**, 3238–3244 (2002).

328. Ose, T. *et al.* Insight into a natural Diels–Alder reaction from the structure of macrophomate synthase. *Nature* **422**, 185–189 (2003).

329. Serafimov, J. M., Gillingham, D., Kuster, S. & Hilvert, D. The Putative Diels−Alderase Macrophomate Synthase is an Efficient Aldolase. *J Am Chem Soc* **130**, 7798–7799 (2008).

330. Guimarães, C. R. W., Udier-Blagović, M. & Jorgensen, W. L. Macrophomate Synthase: QM/MM Simulations Address the Diels−Alder versus Michael−Aldol Reaction Mechanism. *J Am Chem Soc* **127**, 3577–3588 (2005).

331. Kim, H., Ruszczycky, M. W., Choi, S., Liu, Y. & Liu, H. Enzyme-catalysed [4+2] cycloaddition is a key step in the biosynthesis of spinosyn A. *Nature* **473**, 109–12 (2011).

332. Pang, B., Zhong, G., Tang, Z. & Liu, W. Enzymatic [4+2] Cycloadditions in the Biosynthesis of Spirotetramates and Spirotetronates. *Methods in enzymology* **575**, 39–63 (2016).

333. Zheng, Q., Tian, Z. & Liu, W. Recent advances in understanding the enzymatic reactions of [4+2] cycloaddition and spiroketalization. *Curr Opin Chem Biol* **31**, 95–102 (2016).

334. Hashimoto, T. & Kuzuyama, T. Mechanistic insights into Diels-Alder reactions in natural product biosynthesis. *Current Opinion in Chemical Biology* **35**, 117–123 (2016).

335. Oikawa, H. Nature's Strategy for Catalyzing Diels-Alder Reaction. *Cell Chem Biology* **23**, 429–430 (2016).

336. Minami, A. & Oikawa, H. Recent advances of Diels–Alderases involved in natural product biosynthesis. *The Journal of Antibiotics* **69**, 500–506 (2016).

337. Hashimoto, T. *et al.* Biosynthesis of Versipelostatin: Identification of an Enzyme-Catalyzed [4+2]-Cycloaddition Required for Macrocyclization of Spirotetronate-Containing Polyketides. *J Am Chem Soc* **137**, 572–575 (2015).

338. Tian, Z. *et al.* An enzymatic [4+2] cyclization cascade creates the pentacyclic core of pyrroindomycins. *Nature Chemical Biology* **11**, 259–265 (2015).

339. Wang, P., Bashiri, G., Gao, X., Sawaya, M. R. & Tang, Y. Uncovering the Enzymes that Catalyze the Final Steps in Oxytetracycline Biosynthesis. *J Am Chem Soc* **135**, 7138–7141 (2013).

340. Hess, B. A. & Smentek, L. Concerted, highly asynchronous, enzyme-catalyzed [4 + 2] cycloaddition in the biosynthesis of spinosyn A; computational evidence. *Org Biomol Chem* **10**, 7503 (2012).

341. Medvedev, M. G. *et al.* Quantifying Possible Routes for SpnF-Catalyzed Formal Diels-Alder Cycloaddition. *Journal of the American Chemical Society* (2017) doi:10.1021/jacs.6b13243.

342. Jeon, B. *et al.* Investigation of the mechanism of the SpnF-catalyzed [4+2]-cycloaddition reaction in the biosynthesis of spinosyn A. *Proceedings of the National Academy of Sciences* **114**, 10408–10413 (2017).

343. Yang, Z. *et al.* Influence of water and enzyme SpnF on the dynamics and energetics of the ambimodal [6+4]/[4+2] cycloaddition. *Proceedings of the National Academy of Sciences* **115**, E848–E855 (2018).

344. Dan, Q. *et al.* Fungal indole alkaloid biogenesis through evolution of a bifunctional reductase/Diels–Alderase. *Nat Chem* 1–9 (2019) doi:10.1038/s41557-019-0326-6.

345. Patel, S. S., Wong, I. & Johnson, K. A. Pre-steady-state kinetic analysis of processive DNA replication including complete characterization of an exonuclease-deficient mutant. *Biochemistry-us* **30**, 511–525 (1991).

346. Wong, I., Patel, S. S. & Johnson, K. A. An induced-fit kinetic mechanism for DNA replication fidelity: direct measurement by single-turnover kinetics. *Biochemistry-us* **30**, 526–537 (1991).

347. Stivers, J. T., Shuman, S. & Mildvan, A. S. Vaccinia DNA topoisomerase I: Single-turnover and steady-state kinetic analysis of the DNA strand cleavage and ligation reactions. *Biochemistry-us* **33**, 327–339 (1994).

348. Kati, W. M., Johnson, K. A., Jerva, L. F. & Anderson, K. S. Mechanism and fidelity of HIV reverse transcriptase. *J Biological Chem* **267**, 25988–97 (1992).

349. Pavon, J. A., Eser, B., Huynh, M. T. & Fitzpatrick, P. F. Single Turnover Kinetics of Tryptophan Hydroxylase: Evidence for a New Intermediate in the Reaction of the Aromatic Amino Acid Hydroxylases. *Biochemistry-us* **49**, 7563–7571 (2010).

350. Liang, A. D. & Lippard, S. J. Single Turnover Reveals Oxygenated Intermediates in Toluene/o-Xylene Monooxygenase in the Presence of the Native Redox Partners. *J Am Chem Soc* **137**, 10520–3 (2015).

351. Marsh, C. O. *et al.* A Natural Diels-Alder Biocatalyst Enables Efficient [4+2] Cycloaddition Under Harsh Reaction Conditions. *Chemcatchem* **11**, 5027–5031 (2019).

352. Baker, D. What has de novo protein design taught us about protein folding and biophysics? *Protein Sci Publ Protein Soc* **28**, 678–683 (2019).

353. Dou, J. *et al.* De novo design of a fluorescence-activating β-barrel. *Nature* 1–7 (2018) doi:10.1038/s41586-018-0509-0.

354. Stetefeld, J., McKenna, S. A. & Patel, T. R. Dynamic light scattering: a practical guide and applications in biomedical sciences. *Biophysical Rev* **8**, 409–427 (2016).

355. Koppel, D. E. Analysis of Macromolecular Polydispersity in Intensity Correlation Spectroscopy: The Method of Cumulants. *J Chem Phys* **57**, 4814–4820 (1972).

356. Gutfreund, H. Rapid-flow techniques and their contributions to enzymology. *Trends in biochemical sciences* (1999).

357. Brewer, C. B. & Peterson, J. A. Single Turnover Kinetics of the Reaction between Oxycytochrome P-45Ocam and Reduced Putidaredoxin. *JBC* (1988).

358. Sassa, A., Beard, W. A., Shock, D. D. & Wilson, S. H. Steady-state, Pre-steady-state, and Single-turnover Kinetic Measurement for DNA Glycosylase Activity. *J Vis Exp* (2013) doi:10.3791/50695.

359. Pudney, C. R. *et al.* Mutagenesis of Morphinone Reductase Induces Multiple Reactive Configurations and Identifies Potential Ambiguity in Kinetic Analysis of Enzyme Tunneling Mechanisms. *Journal of the American Chemical Society* **129**, 13949–13956 (2007).

360. Ferrer, S. *et al.* A Theoretical Analysis of Rate Constants and Kinetic Isotope Effects Corresponding to Different Reactant Valleys in Lactate Dehydrogenase. *J Am Chem Soc* **128**, 16851–16863 (2006).

361. Gilbert, H. F. Catalysis of thiol/disulfide exchange: single-turnover reduction of protein disulfide-isomerase by glutathione and catalysis of peptide disulfide reduction. *Biochemistry-us* **28**, 7298–7305 (1989).

362. Rivero, U., Meuwly, M. & Willitsch, S. A computational study of the Diels-Alder reactions between 2,3-dibromo-1,3-butadiene and maleic anhydride. *Chem Phys Lett* **683**, 598–605 (2017).

363. Yepes, D. *et al.* The reaction force constant as an indicator of synchronicity/nonsynchronicity in [4+2] cycloaddition processes. *Phys Chem Chem Phys Pccp* **15**, 7311–20 (2013).

364. Hammes, G. G. Multiple Conformational Changes in Enzyme Catalysis †. *Biochemistry-us* **41**, 8221–8228 (2002).

365. Wong, I., Lundquist, A. J., Bernards, A. S. & Mosbaugh, D. W. Presteady-state Analysis of a Single Catalytic Turnover by Escherichia coli Uracil-DNA Glycosylase Reveals a "Pinch- Pull -Push" Mechanism. *J Biol Chem* **277**, 19424–19432 (2002).

366. Wester, M. R. *et al.* Structure of a Substrate Complex of Mammalian Cytochrome P450 2C5 at 2.3 Å Resolution: Evidence for Multiple Substrate Binding Modes † , ‡. *Biochemistry-us* **42**, 6370–6379 (2003).

367. Birdsall, B., Feeney, J., Tendler, S. J. B., Hammond, S. J. & Roberts, G. C. K. Dihydrofolate reductase: multiple conformations and alternative modes of substrate binding. *Biochemistry-us* **28**, 2297–2305 (1989).

368. Shi, J. *et al.* Multiple States of the Tyr318Leu Mutant of Dihydroorotate Dehydrogenase Revealed by Single-Molecule Kinetics. *Journal of the American Chemical Society* **126**, 6914–6922 (2004).

369. Bridgeman, P. W. THE COAGULATION OF ALBUMEN BY PRESSURE. *J. Biol. Chem.* 511–512 (1914).

370. Robb, F. & Clark, D. Adaptation of proteins from hyperthermophiles to high pressure and high temperature. *J. Molec. Microbiol. Biotechnol.* (1999).

371. Mozhaev, V. V., Heremans, K., Frank, J., Masson, P. & Balny, C. High pressure effects on protein structure and function. *Proteins Struct Funct Bioinform* **24**, 81–91 (1996).

372. Kauzmann, W. Thermodynamics of unfolding. *Nature* **325**, 763 (1987).

373. Penniston, J. T. High hydrostatic pressure and enzymic activity: Inhibition of multimeric enzymes by dissociation. *Arch Biochem Biophys* **142**, 322–332 (1971).

374. GROSS, M. & JAENICKE, R. Proteins under pressure. *Eur J Biochem* **221**, 617–630 (1994).

375. Zipp, A. & Kauzmann, W. Pressure denaturation of metmyoglobin. *Biochemistry-us* **12**, 4217–4228 (1973).

376. Luong, T., Kapoor, S. & Winter, R. Pressure—A Gateway to Fundamental Insights into Protein Solvation, Dynamics, and Function. *ChemPhysChem* **16**, 3555–3571 (2015).

377. Czeslik, C., Luong, T. Q. & Winter, R. Enzymatic activity under pressure. *Mrs Bull* **42**, 738–742 (2017).

378. Kremer, W. *et al.* Pulsed pressure perturbations, an extra dimension in NMR spectroscopy of proteins. *Journal of the American Chemical Society* (2011) doi:10.1021/ja2050698.

379. Hay, S. *et al.* Nature of the energy landscape for gated electron transfer in a dynamic redox protein. *Journal of the American Chemical Society* **132**, 9738–45 (2010).

380. Chen, B., Hoffmann, R. & Cammi, R. The Effect of Pressure on Organic Reactions in Fluids—a New Theoretical Perspective. *Angewandte Chemie Int Ed* **56**, 11126–11142 (2018).

381. Schettino, V. & Bini, R. Constraining molecules at the closest approach: chemistry at high pressure. *Chemical Society Reviews* **36**, 869–880 (2007).

382. Taniguchi, Y. & Suzuki, K. Studies of polymer effects under pressure. Part 7. Pressure inactivation of .alpha.-chymotrypsin. *J Phys Chem* **87**, 5185–5193 (1983).

383. Mozhaev, V. V., Lange, R., Kudryashova, E. V. & Balny, C. Application of high hydrostatic pressure for increasing activity and stability of enzymes. *Biotechnol Bioeng* **52**, 320–331 (1996).

384. Akasaka, K., Nagahata, H., Maeno, A. & Sasaki, K. Pressure acceleration of proteolysis: A general mechanism. *BIOPHYSICS* **4**, 29–32 (2008).

385. Luong, T. *et al.* Hydrostatic Pressure Increases the Catalytic Activity of Amyloid Fibril Enzymes. *Angewandte Chemie International Edition* **55**, 12412–12416 (2016).

386. Araujo, T. L. *et al.* Conformational changes in human Hsp70 induced by high hydrostatic pressure produce oligomers with ATPase activity but without chaperone activity. *Biochemistry* **53**, 2884–9 (2014).

387. Eldik, V. R., Asano, T. & Noble, L. W. Activation and reaction volumes in solution. 2. *Chem Rev* **89**, 549–688 (1989).

388. Uroos, M. *et al.* Total synthesis of (−)-aritasone via the ultra-high pressure hetero-Diels–Alder dimerisation of (−)-pinocarvone. *Organic & Biomolecular Chemistry* **15**, 8523–8528 (2017).

389. Butz, P., Greulich, K. O. & Ludwig, H. Volume changes during enzyme reactions: indications of enzyme pulsation during fumarase catalysis. *Biochemistry-us* **27**, 1556–1563 (1988).

390. Jackson, S. E. Ubiquitin: a small protein folding paradigm. *Org Biomol Chem* **4**, 1845 (2006).

391. Eisenmenger, M. J. & Reyes-De-Corcuera, J. I. High pressure enhancement of enzymes: A review. *Enzyme and Microbial Technology* **45**, 331–347 (2009).

392. Morild, E. Advances in Protein Chemistry. *Adv Protein Chem* **34**, 93–166 (1981).

393. Pudney, C. R. *et al.* Parallel Pathways and Free-Energy Landscapes for Enzymatic Hydride Transfer Probed by Hydrostatic Pressure. *ChemBioChem* **10**, 1379–1384 (2009).

394. Hay, S., Pudney, C. R., Sutcliffe, M. J. & Scrutton, N. S. Probing active site geometry using high pressure and secondary isotope effects in an enzyme-catalysed 'deep' H-tunnelling reaction. *Journal of Physical Organic Chemistry* **23**, 696–701 (2010).

395. Kitahara, R., Yamada, H. & Akasaka, K. Two folded conformers of ubiquitin revealed by high-pressure NMR. *Biochemistry* (2001) doi:10.1021/bi010922u.

396. Kitahara, R. & Akasaka, K. Close identity of a pressure-stabilized intermediate with a kinetic intermediate in protein folding. *Proceedings of the National Academy of Sciences* (2003) doi:10.1073/pnas.0630309100.

397. Kachel, N., Kremer, W., Zahn, R. & Kalbitzer, H. R. Observation of intermediate states of the human prion protein by high pressure NMR spectroscopy. *Bmc Struct Biol* **6**, 16 (2006).

398. Kitahara, R., Yokoyama, S. & Akasaka, K. NMR snapshots of a fluctuating protein structure: ubiquitin at 30 bar-3 kbar. *J Mol Biol* **347**, 277–85 (2005).

399. Lin, M.-C., Eid, P., Wong, P. T. T. & Macgregor, R. B. High pressure fourier transform infrared spectroscopy of poly(dA)poly(dT), poly(dA) and poly(dT). *Biophys Chem* **76**, 87–94 (1999).

400. Dzwolak, W., Kato, M. & Taniguchi, Y. Fourier transform infrared spectroscopy in high-pressure studies on proteins. *Biochimica Et Biophysica Acta Bba - Protein Struct Mol Enzym* **1595**, 131–144 (2002).

401. Lerch, M. T., Horwitz, J., McCoy, J. & Hubbell, W. L. Circular dichroism and site-directed spin labeling reveal structural and dynamical features of high-pressure states of myoglobin. *Proceedings of the National Academy of Sciences* **110**, E4714–E4722 (2013).

402. Terefe, N., Buckow, R. & Versteeg, C. Quality-Related Enzymes in Fruit and Vegetable Products: Effects of Novel Food Processing Technologies, Part 1: High-Pressure Processing. *Critical Reviews in Food Science and Nutrition* **54**, 24–63 (2013).

403. Terefe, N., Sheean, P., Fernando, S. & Versteeg, C. The stability of almond β-glucosidase during combined high pressure–thermal processing: a kinetic study. *Applied Microbiology and Biotechnology* **97**, 2917–2928 (2013).

404. Tribst, A., Cota, J., Murakami, M. & Cristianini, M. Effects of High Pressure Homogenization on the Activity, Stability, Kinetics and Three-Dimensional Conformation of a Glucose Oxidase Produced by Aspergillus niger. *PLoS ONE* **9**, e103410 (2014).

405. Micsonai, A. *et al.* Accurate secondary structure prediction and fold recognition for circular dichroism spectroscopy. *Proceedings of the National Academy of Sciences* **112**, E3095–E3103 (2015).

406. Kalbitzer, H., Spoerner, M., Ganser, P., Hozsa, C. & Kremer, W. Fundamental link between folding states and functional states of proteins. *Journal of the American Chemical Society* (2009) doi:10.1021/ja904314q.

407. Roche, J. *et al.* Cavities determine the pressure unfolding of proteins. *P Natl Acad Sci Usa* (2012) doi:10.1073/pnas.1200915109.

408. Tang, L. *et al.* Cloning and Heterologous Expression of the Epothilone Gene Cluster. *Science* **287**, 640–642 (2000).

409. Aldrich, C. C., Beck, B. J., Fecik, R. A. & Sherman, D. H. Biochemical Investigation of Pikromycin Biosynthesis Employing Native Penta- and Hexaketide Chain Elongation Intermediates. *J Am Chem Soc* **127**, 8441–8452 (2005).

# Appendices

## A.1 Primers and Plasmids Used in This Project

### A.1.1 Primers Used

| Primer Name | Protein | Primer Sequence 5'-3' | Primer function |
|---|---|---|---|
| B2_pET28_Fwd | AbyB2 | tgggtcgcggatccgaattcATGAC-CACACCGAGCAAAGG | Insertion into pET28 |
| B2_pET28_Rvs | AbyB2 | caagcttgtcgacggagctcTTAAC-CTGCGTCAACTTCTTGATAATC | Insertion into pET28 |
| B2_pF_Fwd (B2_1F) | AbyB2 | aagttctgtttcagggcccgATGAC-CACACCGAGCAAAGG | Insertion into pOPINF, sequencing |
| B2_pF_Rvs (B2_6R) | AbyB2 | atggtctagaaagctttaTTAAC-CTGCGTCAACTTCTTGATAATC | Insertion into pOPINF, sequencing |
| B2_1CF | AbyB2 | GGCAGAATATGGTCAGGGTC | Sequencing & assembly verification |
| B2_1CR | AbyB2 | AACAGTTCCGGTGTGGC | Sequencing |
| B2_1R | AbyB2 | ACGCAGTGACCAATCCACATG | Sequencing |
| B2_2F | AbyB2 | GATTGGTCACTGCGTGCAG | Sequencing |
| B2_2CF | AbyB2 | GTCTGGGTGCCGTTGG | Sequencing & assembly verification |
| B2_2CR | AbyB2 | ACGACGACGATCACCTTC | Sequencing & assembly verification |
| B2_2R | AbyB2 | ACCATCAGGTTCTGCCAGAC | Sequencing |
| B2_3F | AbyB2 | GCAGAACCTGATGGTCCGC | Sequencing |
| B2_3CF | AbyB2 | GCCTGACCGCAATTGAAC | Sequencing & assembly verification |
| B2_3CR | AbyB2 | GCTCACCAGACCTAAACGA | Sequencing & assembly verification |
| B2_3R | AbyB2 | GCTCAGAACGGTGGCACC | Sequencing |
| B2_4F | AbyB2 | GCCACCGTTCTGAGCAATC | Sequencing |
| B2_4CF | AbyB2 | GAAGAAGCCCTGCGTCG | Sequencing |
| B2_4CR | AbyB2 | GTGCTGCCGGATTTGCT | Sequencing & assembly verification |
| B2_4R | AbyB2 | CAGTGCTGTCAGAAAACGACG | Sequencing |
| B2_5F | AbyB2 | TTTCTGACAGCACTGGCCG | Sequencing |
| B2_5CF | AbyB2 | GGTTACCAGCACCATTGC | Sequencing & assembly verification |
| B2_5CR | AbyB2 | CCAAATGCCAGTGCATGC | Sequencing |

| | | | |
|---|---|---|---|
| B2_5R | AbyB2 | AATCAGAACACGTTCACCTGGAC | Sequencing |
| B2_6F | AbyB2 | GAACGTGTTCTGATTCATAGCGC | Sequencing |
| B2_6CF | AbyB2 | GGGTCGTATTGAACGTGGT | Sequencing |
| B2_6CR | AbyB2 | GGAACGGCTGCCAGAAC | Sequencing & assembly verification |
| abyB2_frag1_FW | AbyB2 | CCGGTCGGTCTCTCTGGATGACCACACCGAGCAAAG | Type IIs assembly |
| abyB2_frag1_RV | AbyB2 | AGGGCAGGTCTCGAGTGACCAATCCACATGAGGAGC | Type IIs assembly |
| abyB2_frag2_FW | AbyB2 | CGACTGGGTCTCCCACTGCGTGCAGTTCTGGCA | Type IIs assembly |
| abyB2_frag2_RV | AbyB2 | AACAATGGTCTCCCGGACCATCAGGTTCTGCCAGA | Type IIs assembly |
| abyB2_frag3_FW | AbyB2 | GCCAACGGTCTCTTCCGCGTCAGGTTGTTGTTGCAC | Type IIs assembly |
| abyB2_frag3_RV | AbyB2 | TCGACCGGTCTCTTGCTCAGAACGGTGGCACCA | Type IIs assembly |
| abyB2_frag4_FW | AbyB2 | TCCAAAGGTCTCCAGCAATCCGGATATTTTTGTTGGTTTT | Type IIs assembly |
| abyB2_frag4_RV | AbyB2 | TCCTGAGGTCTCCGCCAGTGCTGTCAGAAAACGACGA | Type IIs assembly |
| abyB2_frag5_FW | AbyB2 | TCACCGGGTCTCGTGGCCGAAGCATATGTGACCGGT | Type IIs assembly |
| abyB2_frag5_RV | AbyB2 | TGCCGCGGTCTCCGAACACGTTCACCTGGACGCA | Type IIs assembly |
| abyB2_frag6_FW | AbyB2 | CATTTCGGTCTCCGTTCTGATTCATAGCGCAGCCG | Type IIs assembly |
| abyB2_frag6_RV | AbyB2 | TTCGGCGGTCTCCATGCTTAACCTGCGTCAACTTCTTGAT | Type IIs assembly |
| AbyB3_pE_Fwd | AbyB3 / AbyB3ΔACP | aggagatataccatgGTGAGCGAGA-CACGC | Insertion into pOPINE, sequencing |
| AbyB3_pE_RVS | AbyB3 | gtgatggtgatgtttTCACGTCCGCGCG | Insertion into pOPINE, sequencing |
| AbyB3no-ACP_pE_Rvs | AbyB3ΔACP | gtgatggtgatgtttCTCGTGTGGCGTCGA | Insertion into pOPINE, sequencing |
| AbyB3_pF_Fwd | AbyB3 / AbyB3ΔACP | aagttctgtttcagggcccgATGAGCGAGACACGCGAGGAGAAG | Insertion into pOPINF, sequencing |
| AbyB3_pF_Rvs | AbyB3 | atggtctagaaagctttaTCAC-GTCCGCGCG | Insertion into pOPINF, sequencing |
| AbyB3no-ACP_pF_Rvs | AbyB3ΔACP | atggtctagaaa-gctttaCTCGTGTGGCGTCGA | Insertion into pOPINF, sequencing |

| | | | |
|---|---|---|---|
| B3_F1_seq | AbyB3 / AbyB3Δ ACP | CTCGTCGTCGCTGGT | Sequencing |
| B3_F2_seq | AbyB3 / AbyB3Δ ACP | CGTCGACGAGCCCAC | Sequencing |
| B3_F3_seq | AbyB3 / AbyB3Δ ACP | GCGTTGTTCACCCTGGA | Sequencing |
| B3_F4_seq | AbyB3 | GCCGTTTCGTGGAGGTC | Sequencing |
| AbyU_pF_Fwd | AbyU | aagttctgtttcagggcccgATGACTGAGC GACTGGAG | Insertion into pOPINF |
| AbyU_pF_Rvs | AbyU | atggtctagaaagctttaTCAC-TCGCCGAGCAG | Insertion into pOPINF |
| T7 (Eurofins) | | TAA TAC GAC TCA CTA TAG GG | Sequencing |
| T7term (Eurofins) | | CTA GTT ATT GCT CAG CGG T | Sequencing |

**Table A.1 List of primers used in this project**

Each primer was designed on Benchling™ to have a Tm as close to 62ºC as possible under the conditions used by CloneAmp HiFi PCR premix (Takara, Japan), apart from GeneArt type IIs assembly primers, which were designed using the Type IIs primer design tool (Thermofisher UK). Lowercase letters encode plasmid overlap sequences for insertion *via* homologous recombination. *Bsa*I recognition sites are highlighted in grey.

### A.1.2 Plasmids Used

| Plasmid | Resistance Marker | Promoter | Tag Sequence | Tag Location |
|---|---|---|---|---|
| pOPINF | Ampicillin | T7 | MAHHHHHHSSGLEVLFQGP | N-terminal |
| pOPINE | Ampicillin | T7 | KHHHHHH | C-terminal |
| pET28a | Kanamycin | T7 | GSSHHHHHHSSGLVPR | N-terminal |

**Table A.2 Plasmids Used in the Project**

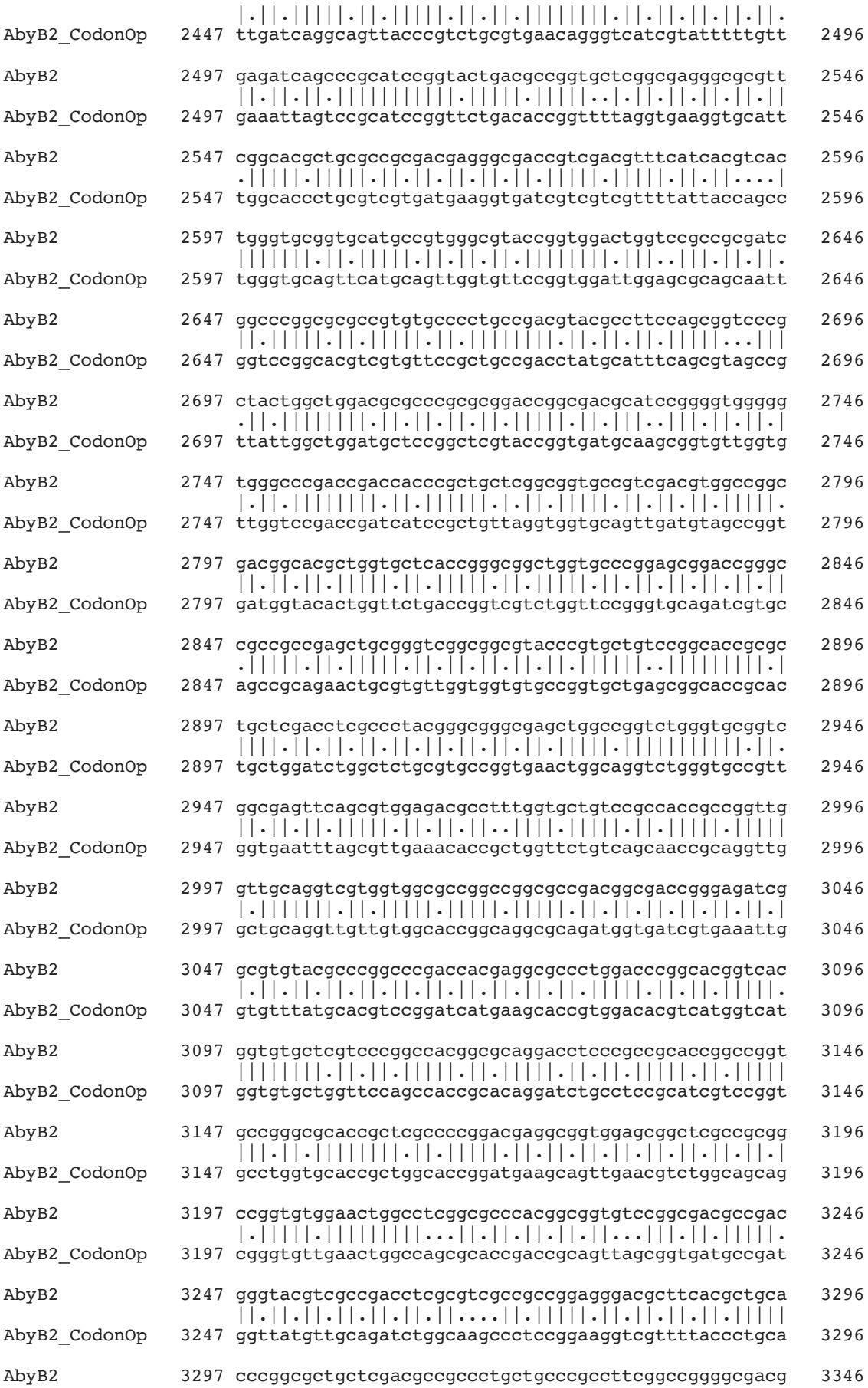
## A.2 Nucleotide and Amino Acid Sequences

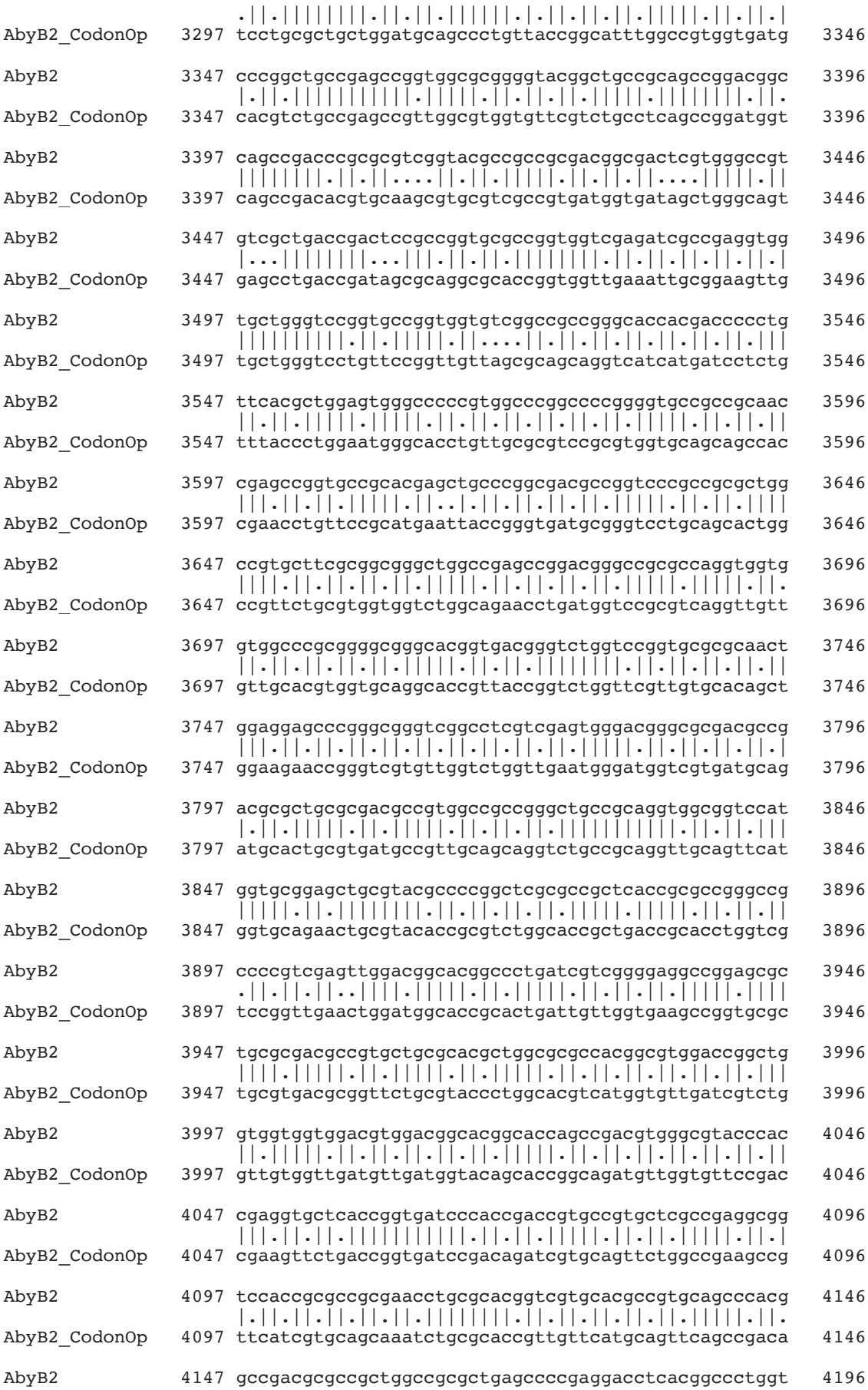
### A.2.1 A Comparison of the Native and Codon Optimised AbyB2 Genes

```
AbyB2              1 atgaccacaccgagcaaggggacggacagcattgccgaccagcagaagct     50
                     ||||||||||||||||||.||.||.||||||||.||.|||||||||.||
AbyB2_CodonOp      1 atgaccacaccgagcaaaggcaccgatagcattgcagatcagcagaaact     50

AbyB2             51 gcgcgagtacctccggcgggtcaccgacgacctgctgcggacccgccggc    100
                     |||.||.||.||.||.||.||.|||||.||.||||||||.|||||.||.|
AbyB2_CodonOp     51 gcgtgaatatctgcgtcgtgttaccgatgatctgctgcgtacccgtcgtc    100

AbyB2            101 gcctgaccgaggtggagtcggccgaccgggagccggtggccatcgtctcg    150
                     |.|||||||||.||.||....||.||.||.||.|||||.||.||.||....
AbyB2_CodonOp    101 gtctgaccgaagttgaaagcgcagatcgtgaaccggttgcaattgttagc    150

AbyB2            151 atggcctgccgcttccccggcggggtggcttcgccggaggacctgtggca    200
                     |||||.||.||.||.||.||.||.||....||||||.||.|||||||||
AbyB2_CodonOp    151 atggcatgtcgttttccaggtggtgttgcaagtccggaagatctgtggca    200

AbyB2            201 gctggtggcgtccggcaccgacgcgatcagcgggttccccgatgaccggg    250
                     |||||||.||...|||||||||.||.||.|||||.||.||.|||||.||.|
AbyB2_CodonOp    201 gctggttgcaagcggcaccgatgcaattagcggttttccggatgatcgtg    250

AbyB2            251 gctggccactggacgagctgtacgacccggacccggagcacccgggcacg    300
                     |.||||||.|||||.||.||||||.||.||.||.|||||.||.|||||.||.
AbyB2_CodonOp    251 gttggcctctggatgaactgtatgatcctgatccggaacatccgggtaca    300

AbyB2            301 tccaccacccggcaggggggtttcctgcacgacgccgccgacttcgaccc    350
                     ..||||||.||.||.||.|||||.|||||.||.||.||.||.||.|||||
AbyB2_CodonOp    301 agcaccacacgtcaaggtggttttctgcatgatgcagcagattttgaccc    350

AbyB2            351 ggagttcttcgggatcagcccccgtgaggcgttgaccatcgacccgcagc    400
                     |||.||.||.||.||.||.||.|||||.||..|||||||.||.|||||||
AbyB2_CodonOp    351 ggaatttttggtattagtccgcgtgaagcactgaccattgatccgcagc    400

AbyB2            401 agcggctgctgttggagaccgcctgggaggcggtggagcgggccgggatc    450
                     ||||.||||||||.||.|||||.|||||.||.||.||.||.||.||.||.
AbyB2_CodonOp    401 agcgtctgctgttagaaaccgcatgggaagcagttgaacgtgcaggtatt    450

AbyB2            451 gcaccggactcgctgcgtggcagccggaccggcgtgttcgccggggtcat    500
                     ||||||||....||||||||.|||||.|||||.||.||.|||||.||.||
AbyB2_CodonOp    451 gcaccggatagcctgcgtggtagccgtaccggtgtttttgccggtgttat    500

AbyB2            501 gtacggcgactacggtgcccgcctgcgcccgatcccggcgggcttcgagg    550
                     |||.||.||.||.|||||.||.|||||.|||||.|||||.||.||.||.|
AbyB2_CodonOp    501 gtatggtgattatggtgcacgtctgcgtccgattccggcaggttttgaag    550

AbyB2            551 ggtacatgggcaccggcagcgcgggcagcgtggccaccggccggatcgcg    600
                     |.||.||||||||||||.|||||.||.|||||.||||||||.||.||.||.
AbyB2_CodonOp    551 gttatatgggcaccggtagcgcaggtagcgttgccaccggtcgtattgca    600

AbyB2            601 tacaccctcggcctggaggggccggcggtgagcgtcgacacggcgtgct-    649
                     ||.|||||.||..|.||.||.|||||.||.|||||.||.||.||.||..
AbyB2_CodonOp    601 tataccctgggtttagaaggtccggcagttagcgttgataccgcatgtag    650

AbyB2            650 cgtcgtcgctggtggcgctgcacctcgcggcgcaggcgctgcggcgcggc    699
                     |..|..| |||||.||.|||||.||.||.||.|||||.|||||.|||||.
AbyB2_CodonOp    651 cagcagc-ctggttgccctgcatctggcagcacaggcactgcgtcgcggt    699

AbyB2            700 gagtgcgacctggcgctggccggcggcgtgacggtcatcgccacgccgga    749
                     ||.||.||.|||||.|||||.|||||.||.||.||.||.|||||.|||||
AbyB2_CodonOp    700 gaatgtgatctggcactggcaggcggtgttaccgttattgccacaccgga    749

AbyB2            750 gctgttcgtggagttcagccgccagcgggggttgtcgccggacgggcggt    799
```

```
                    .|||||.||.||.||.|||||.|||||.||..||...|||||.||.||.|
AbyB2_CodonOp    750 actgtttgttgaatttagccgtcagcgtggtctgagtccggatggtcgtt    799

AbyB2            800 gcaaggcattcgcggcctcggccgacggcacgggctgggccgagggcgtc    849
                     |.||.|||||.||.|||...||.||.||.||.||.|||||.||.||.||.
AbyB2_CodonOp    800 gtaaagcatttgcagccagcgcagatggtacaggttgggcagaaggtgtt    849

AbyB2            850 ggcctggtgcttgtcgaacggctcgccgacgcccgccgcaacggtcaccc    899
                     ||.|||||.||.||.|||||.||.||.||.||.||.||.|||||.||
AbyB2_CodonOp    850 ggtctggttctggttgaacgtctggcagatgcacgtcgtaatggtcatcc    899

AbyB2            900 ggtgctcgcgctgctgcgcggcagcgccgtcaaccaggacggccgcagca    949
                     |||.||.||.|||||||.||....||.||.||.|||||.||.|||||||
AbyB2_CodonOp    900 ggttctggccctgctgcgtggttcagcagttaatcaggatggtcgcagca    949

AbyB2            950 gccagctctccgcgcccaacggcccggcgcagcggcgggtcatccgggcg    999
                     |||||||...|||.||.||.||.|||||.|||||.||.||.||.||.||.
AbyB2_CodonOp    950 gccagctgagcgcaccgaatggtccggcacagcgtcgtgtgattcgtgca    999

AbyB2           1000 gcgctggccagtgccggcctggagcccgccgaggtcgacctggtggaggc   1049
                     ||.|||||.||.||.||.|||||.||.||.||.||.||.|||||.||.||
AbyB2_CodonOp   1000 gcactggcaagcgcaggtctggaaccggcagaagttgatctggtagaagc   1049

AbyB2           1050 acacggcaccggcacccggctcggtgacccgatcgaggcccaggcactgc   1099
                     |||.||.||.|||||||.||.|||||.|||||.||.|||||||.||||
AbyB2_CodonOp   1050 acatggtacgggcacccgtctgggtgatccgattgaagcccaggcgctgc   1099

AbyB2           1100 tggccgagtacgggcaggggcgcacggagccgctctggctgggctcgctc   1149
                     ||||.||.||.||.|||||.||.||.||.|||||.|||||||| |.||..
AbyB2_CodonOp   1100 tggcagaatatggtcagggtcgtaccgaaccgctgtggctggg-tagcct   1148

AbyB2           1150 aagtcg-aacatcgggcacacccaggcggcggccggcgtcggcggggtga   1198
                     .|...| ||.||.||.||.|||||||.||.|||||.||.||.||.||.|
AbyB2_CodonOp   1149 gaaaagcaatattggtcatacccaggcagcagccggtgttggtggtgtta   1198

AbyB2           1199 tcaaggtggtgcaggccatgcggcacgggctcctgcccgccaccctgcac   1248
                     |.||.||.||.|||||.|||||.||.||.||.|||||.||.||.|||||.
AbyB2_CodonOp   1199 ttaaagttgttcaggcaatgcgtcatggtctgctgccagcaacactgcat   1248

AbyB2           1249 gccgacgaggccaccccgcacgtcgactggagcgtcggtgacgtccgtct   1298
                     |||||.||.|||||.|||||.||.||.|||||||.|||||.||.|||||
AbyB2_CodonOp   1249 gccgatgaagccacaccgcatgttgattggagcgttggtgatgttcgtct   1298

AbyB2           1299 gctcaccgaggcccgcgactggccggcgcgggagcggccccggcgggcgg   1348
                     |||.||.||.||.||.||.|||||.||.||.||.||.||.||.||.|
AbyB2_CodonOp   1299 gctgacggaagcacgtgattggcctgcacgtgaacgtccgcgtcgtgcag   1348

AbyB2           1349 ccgtgtcctccttcggcatcagcggcaccaacgcgcacgtgatcctggag   1398
                     ||||...|..|||.||.||....|||||||.||.||.||.||.|||||.
AbyB2_CodonOp   1349 ccgttagcagctttggtatttcaggcaccaatgcacatgttattctggaa   1398

AbyB2           1399 gagggcgaccccgacggggtcgccgacgcgccaccggacgacgtgctcgc   1448
                     ||||||||.||.||.||.||.|||||.||.||.||.||.||.||.||.||
AbyB2_CodonOp   1399 gagggcgatccggatggcgtggccgatgcaccgcctgatgatgttctggc   1448

AbyB2           1449 gcgcaagccggtgccggtggtgctgtcggcgcacaccgcgtcggctctcg   1498
                     .||.||.|||||.|||||.||.|||...||.||.|||||....||.|||.|
AbyB2_CodonOp   1449 acgtaaaccggttccggttgttctgagtgcacataccgcaagcgcactgg   1498

AbyB2           1499 ggccgcaggcggcccggctccgcgcgcacctcgacgcccaccccgacctc   1548
                     |.|||||||.|||||.||.||.||.||.||.||.||.||.||.||.||.
AbyB2_CodonOp   1499 gtccgcaggcagcccgtctgcgtgcacatctggatgcacatcctgatctg   1548

AbyB2           1549 acggtggccgacgtcgcgcactcgctggccaccacccgtaccccgctcgc   1598
                     ||.||.||.||.||.||.||....|||||.|||||.|||||.|||||.||
AbyB2_CodonOp   1549 accgttgcagatgttgcacatagcctggcaaccacacgtacaccgctggc   1598

AbyB2           1599 cgagcgggccgtcctcgttgccgccgacctcgacgagctgcgcaccgccc   1648
```

```
                    |||.||.|||||.||.||.||.|||||.||.|||||.|||||.|||||.|
AbyB2_CodonOp  1599 cgaacgtgccgttctggtggcagccgatctggacgaactgcgtaccgcac  1648

AbyB2          1649 tggacgccgtcgcggacggcgacgagccgccggtgcgcggcacggccggg  1698
                    ||||.||.||.||.||.||.||.|||||.||.||.|||||.||.||.
AbyB2_CodonOp  1649 tggatgcagttgccgatggtgatgaaccgcctgttcgtggcaccgcaggt  1698

AbyB2          1699 caccccggcggggtcgtgttcgtcttccccggccagggcgcccagtgggc  1748
                    ||.||.||.||.||.||.||.||.||.||||||||.||.||||||||
AbyB2_CodonOp  1699 catcctggtggtgtggtttttgtttttccaggccagggtgcacagtgggc  1748

AbyB2          1749 gggcatggccctggacctgtaccgcgaggacgaggtgttccgcgcggcgc  1798
                    .||.||||||||||||.|||||.||.||.||.||.||.|||||.||.|
AbyB2_CodonOp  1749 aggtatggccctggatctgtatcgtgaagatgaagttttcgcgcagccc  1798

AbyB2          1799 tggacgactgcgagcgggccctggccccgcacgtcgactggtcgctgcgg  1848
                    ||||.||.||.||.||||||||.||.||.||.||.|||||.|||||.
AbyB2_CodonOp  1799 tggatgattgtgaacgtgccctggctcctcatgtggattggtcactgcgt  1848

AbyB2          1849 gccgtgctggccgacgccgacgccctcggacgcgtcgacgtcgtccagcc  1898
                    ||.||.|||||.||.||.||.||.||.||.||.||.||.||.|||||
AbyB2_CodonOp  1849 gcagttctggcagatgcagatgcactgggtcgtgttgatgttgttcagcc  1898

AbyB2          1899 cgccctgtgggcggtgatggtgtcactggcggcgctgtggcagcaccacg  1948
                    .||.||||||||.||.|||||....||||||.||.||||||||||||.||.|
AbyB2_CodonOp  1899 tgcactgtgggcagttatggttagcctggcagcactgtggcagcatcatg  1948

AbyB2          1949 gcgtgacccccgacgcggtcgtcggccactcccagggcgagatcgccgcc  1998
                    |.||.||.||.||.||.||.||.||...|||||||.||.||.||.||.
AbyB2_CodonOp  1949 gtgttacaccggatgcagttgttggtcatagccagggtgaaattgcagca  1998

AbyB2          1999 gcctgtgtcgcgggtgccctgtcgctggagcaggccgcggccgtggtggc  2048
                    ||.|||||.||.|||||.|||...||||||.|||||.||.|||||.||.||
AbyB2_CodonOp  1999 gcatgtgttgccggtgcactgagcctggaacaggcagcagccgttgttgc  2048

AbyB2          2049 gctgcgggcgcgggccatcaccgccgtcgcgggtcgcggtgccatggcgt  2098
                    .|||||.||.||.|||||.|||||.||.||.|||||.|||||.|||||..
AbyB2_CodonOp  2049 cctgcgtgcacgtgccattaccgcagttgcaggtcgtggtgcaatggcaa  2098

AbyB2          2099 cggtttctgtgcccgcacagcagatcaccgagcggtggggcgaccggatc  2148
                    ..|||...||.||.||||||||||.|||||.||.|||||.||.||.||.
AbyB2_CodonOp  2099 gcgttagcgttccggcacagcagattaccgaacgttggggtgatcgtatt  2148

AbyB2          2149 accgtggccgtgacgaactccgccgacgcgacggtcgtcgccggtgaacc  2198
                    |||||.||.||.||.||...|||.||.||.||.||.||.||.||||||||
AbyB2_CodonOp  2149 accgttgcagttaccaatagcgcagatgcaaccgtggttgcgggtgaacc  2198

AbyB2          2199 agaggccgtcgccgaggtggtcgccgcgtacgacgccgagggagtccggg  2248
                    .||.|||||.|||||.||.||.||.||.||.|||||.||.||.|||.|
AbyB2_CodonOp  2199 ggaagccgttgccgaagttgttgcagcatatgatgccgaaggtgttcgtg  2248

AbyB2          2249 cccgggtcctgccggtggactacgcctcccactcggcgcacgtg--gagc  2296
                    ||||.||.||||||||.||.||.||...|||  ..|||||||.||  ||.|
AbyB2_CodonOp  2249 cccgtgttctgccggttgattatgcaagcca--tagcgcacatgttgaac  2296

AbyB2          2297 ccgtgcgggagccgatcctcgacgcgctgcgcgacctcaccccgaccgag  2346
                    |.||||||.||.||||||.||.||.||.|||||.||.||.||||||||||.
AbyB2_CodonOp  2297 cggtgcgtgaaccgattctggatgcactgcgtgatctgaccccgaccgaa  2346

AbyB2          2347 gcccgggtcccgttccactccaccgtcaccggtgcggagttcgacacccg  2396
                    ||.||.||.|||||.||...|||||||.||||||||.||.||.||.|||||
AbyB2_CodonOp  2347 gcacgtgttccgtttcatagcaccgttaccggtgcagaatttgatacccg  2396

AbyB2          2397 gggcctgaccgccgactactggtacaccaacctgcgcagcaccgtccggt  2446
                    .||.||||||||.||.||.|||||.||.||.|||||.||||||||.||.|
AbyB2_CodonOp  2397 tggtctgaccgcagattattggtatacaaatctgcgtagcaccgtgcgtt  2446

AbyB2          2447 tcgaccaggcggtgacccggctacgcgaacagggccaccggatcttcgtc  2496
```

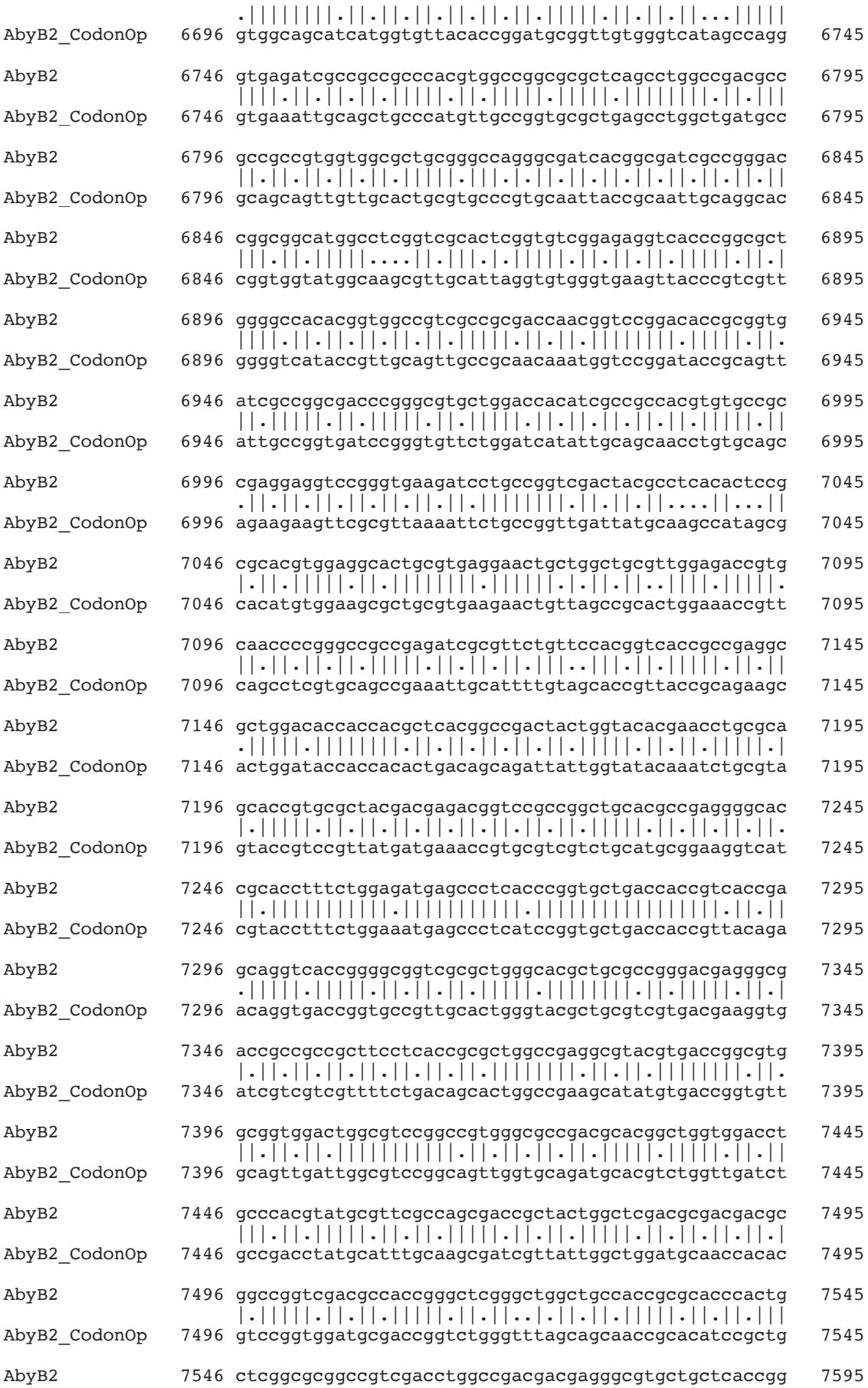


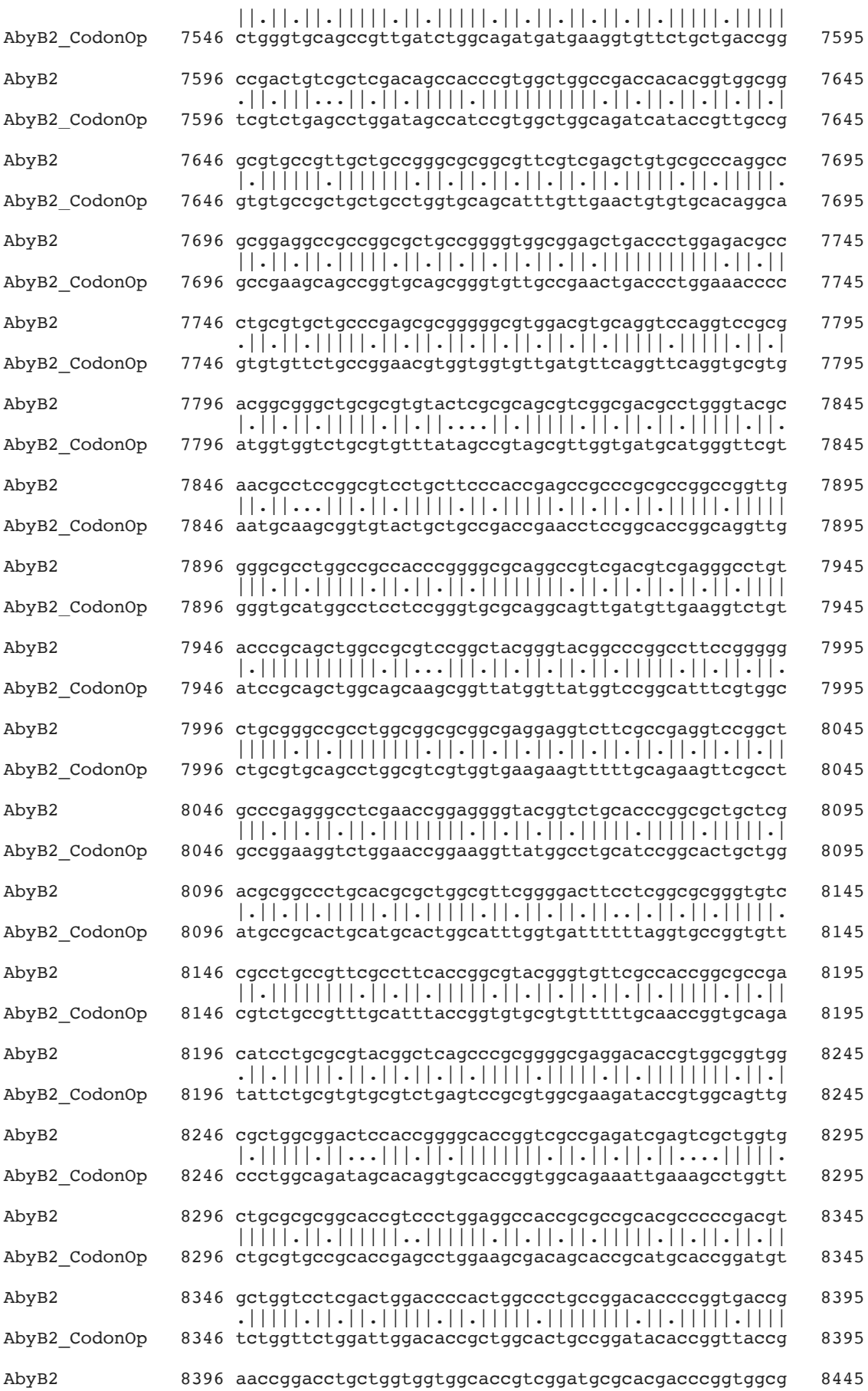


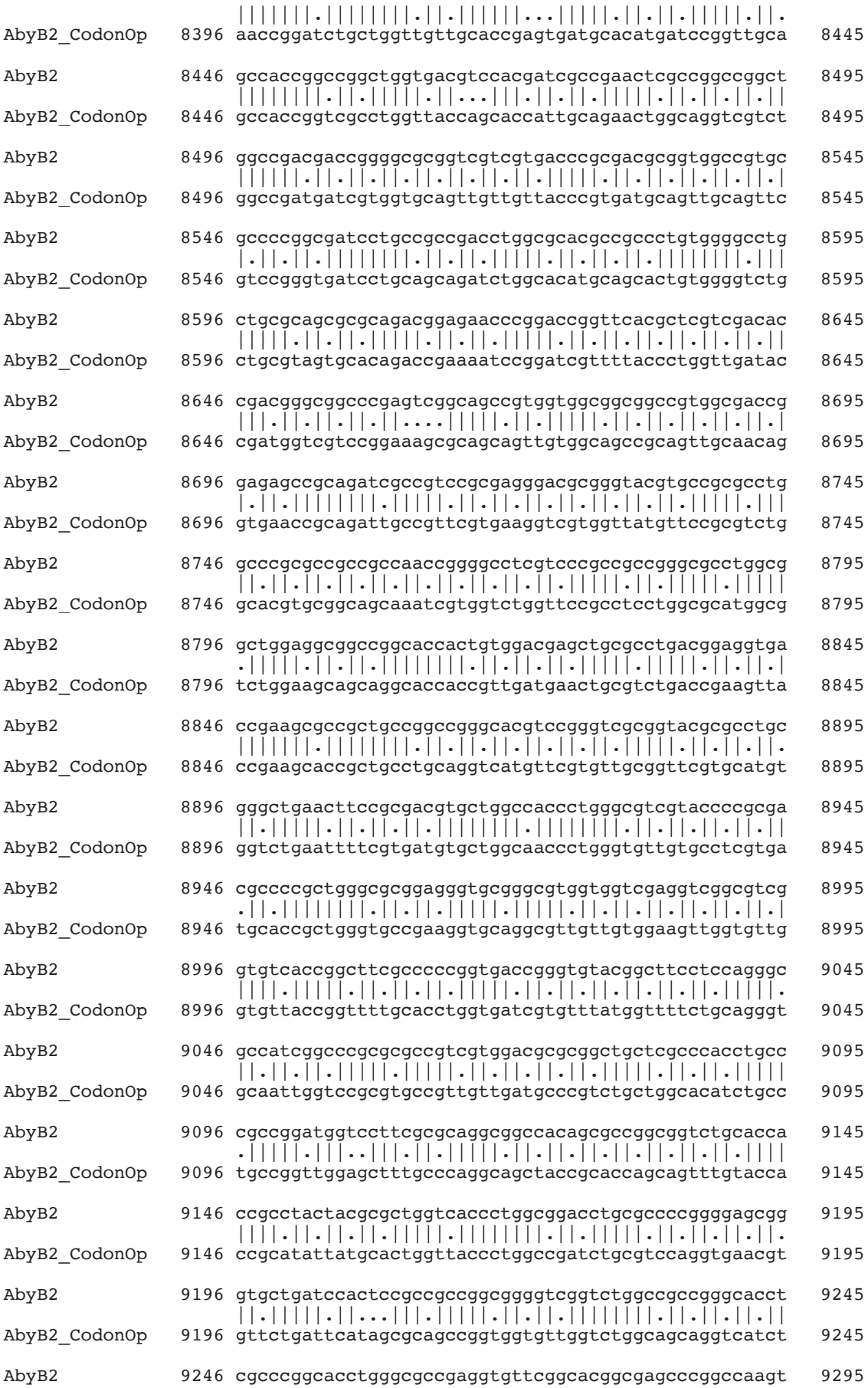
```
                      |||||.||.||||||||.||.|||||.||.||.||.||.||.||.|||||
AbyB2_CodonOp    4147 gccgatgcaccgctggcagcactgagtccggaagatctgacagcactggt   4196

AbyB2            4197 cgagcggatcgtggcgccggcccggcacctgcacgaactgaccgcgcacc   4246
                      .||.||.||.||.||.|||||.||.||.|||||||||||||||.||.||.|
AbyB2_CodonOp    4197 tgaacgtattgttgcaccggcacgccatctgcacgaactgacagcccatc   4246

AbyB2            4247 tgccgctgacccggttcgtgctgtccggctcggcggcgggtgtgctgggc   4296
                      ||||||||||.||.||.||.|||..|||....||.||.|||||..|.||.
AbyB2_CodonOp    4247 tgccgctgacacgttttgttctgagcggtagcgcagccggtgttttaggt   4296

AbyB2            4297 ggcatcgggcaggccgccgtggcggcggccaccaccgggttgggcgcgct   4346
                      ||.||.||.|||||.||.||.||.||.||.|||||.||..||||.||.||
AbyB2_CodonOp    4297 ggtattggtcaggcagcagttgcagccgcaaccacaggtctgggtgcact   4346

AbyB2            4347 cgcggcccgccgtcgcgccgacgggctgcccgcccaggtcgtcgcctggg   4396
                      .||.|||||.|||||.|||||.||.||||||.||.||||||.||.||.||||
AbyB2_CodonOp    4347 ggcagcccgtcgtcgtgccgatggtctgcctgcacaggttgtggcatggg   4396

AbyB2            4397 gtccgtcggccggcacccgccggctcgggctggtgtccctcgacgacgcc   4446
                      |||||...||.||||||||.||..|.||.|||||||..|||.||.||.||.
AbyB2_CodonOp    4397 gtccgagcgcaggcacccgtcgtttaggtctggtgagcctggatgatgct   4446

AbyB2            4447 cgcctggccgcactgttcgcgcaggtcttggcgcacgacgtcgacgtggt   4496
                      ||||||||.||.|||||.||.||||||..|||||.||.||.||.||.||.||
AbyB2_CodonOp    4447 cgcctggcagcgctgtttgcccaggttctggcacatgatgtggatgttgt   4496

AbyB2            4497 ggccgcgcccctggtccgggcggggctgcgggggcaggcgcgggcgggca   4546
                      .|||||.||.|||||.||.||.||.|||||.||.||.||.||.||.|||.|
AbyB2_CodonOp    4497 tgccgctccgctggttcgtgccggtctgcgtggtcaagcacgtgcaggta   4546

AbyB2            4547 cgctgcccgtggcgctgcgggccctcgtgccggcgctgcccggcggtgcc   4596
                      |.|||||.||.||.|||||.||.||.||.|||||.|||||.||.|||||.
AbyB2_CodonOp    4547 cactgccggttgctctgcgtgctctggttccggcactgcctggtggtgca   4596

AbyB2            4597 accggcctggcggcccggctggccggcgcctcccccgccgaggggcgccg   4646
                      ||.||..|.||.||.||.|||||.|||||....||.||.||.||.||.||
AbyB2_CodonOp    4597 actggtttagcagcacgtctggcaggcgcaagtccggcagaaggtcgtcg   4646

AbyB2            4647 tctgctgctggacaccatccgtacccatgtcgccggggtgctgggccacg   4696
                      ||||||||||||.|||||.||||||||||||.|||||.||.|||||.||.|
AbyB2_CodonOp    4647 tctgctgctggataccattcgtacccatgttgccggtgttctgggtcatg   4696

AbyB2            4697 acgacgccagcggcatcgacgagcgccgcgccttcaaggacctcgggttc   4746
                      |.||.||.|||||.||.||.||.||.||.||.||.||.||.||.||.||.
AbyB2_CodonOp    4697 atgatgcaagcggtattgatgaacgtcgtgcatttaaagatctgggtttt   4746

AbyB2            4747 gactcgctcaccgccatcgagctgcgcaaccgtctcaacaccgcgctggg   4796
                      ||....||.|||||.||.||.||||||||.|||||.||.|||||.|||||
AbyB2_CodonOp    4747 gatagcctgaccgcaattgaactgcgcaatcgtctgaataccgcactggg   4796

AbyB2            4797 gcggaccctgcccgcgacgctgatcttcgaccaccccagtcccggcgctc   4846
                      .||.||||||||.||.||.|||||.||.||.||.||.|||||.||.||.|
AbyB2_CodonOp    4797 tcgtaccctgcctgcaacactgatttttgatcatccgagtccgggtgcgc   4846

AbyB2            4847 tggccgaacacctgcgcgacgacctgctcggccgtgccgccgtggccgct   4896
                      ||||||||||.|||||.||.||.|||||.||.|||||.|||||.|||||.
AbyB2_CodonOp    4847 tggccgaacatctgcgggatgatctgctgggtcgtgcagccgttgccgca   4896

AbyB2            4897 gccccggtggcggtcgccagcgacgaaccgatcgccatcgtcgcgatggg   4946
                      ||.||||||.|||||.||....||.||||||||.||.||.||.||.|||||
AbyB2_CodonOp    4897 gcaccggttgcggttgcatcagatgaaccgattgcaattgttgcaatggg   4946

AbyB2            4947 ctgccgctatccgggcggcatcgccgaccccgaggcgttgtggcaggcgg   4996
                      .||.||.|||||.||.|||||.||.||.||.||.||..||||||||||.|
AbyB2_CodonOp    4947 ttgtcgttatcctggtggcattgcagatccggaagcactgtggcaggcag   4996

AbyB2            4997 tggtgtcggaactcgacgcggtgggaccgttcccgaccgaccggggctgg   5046
```

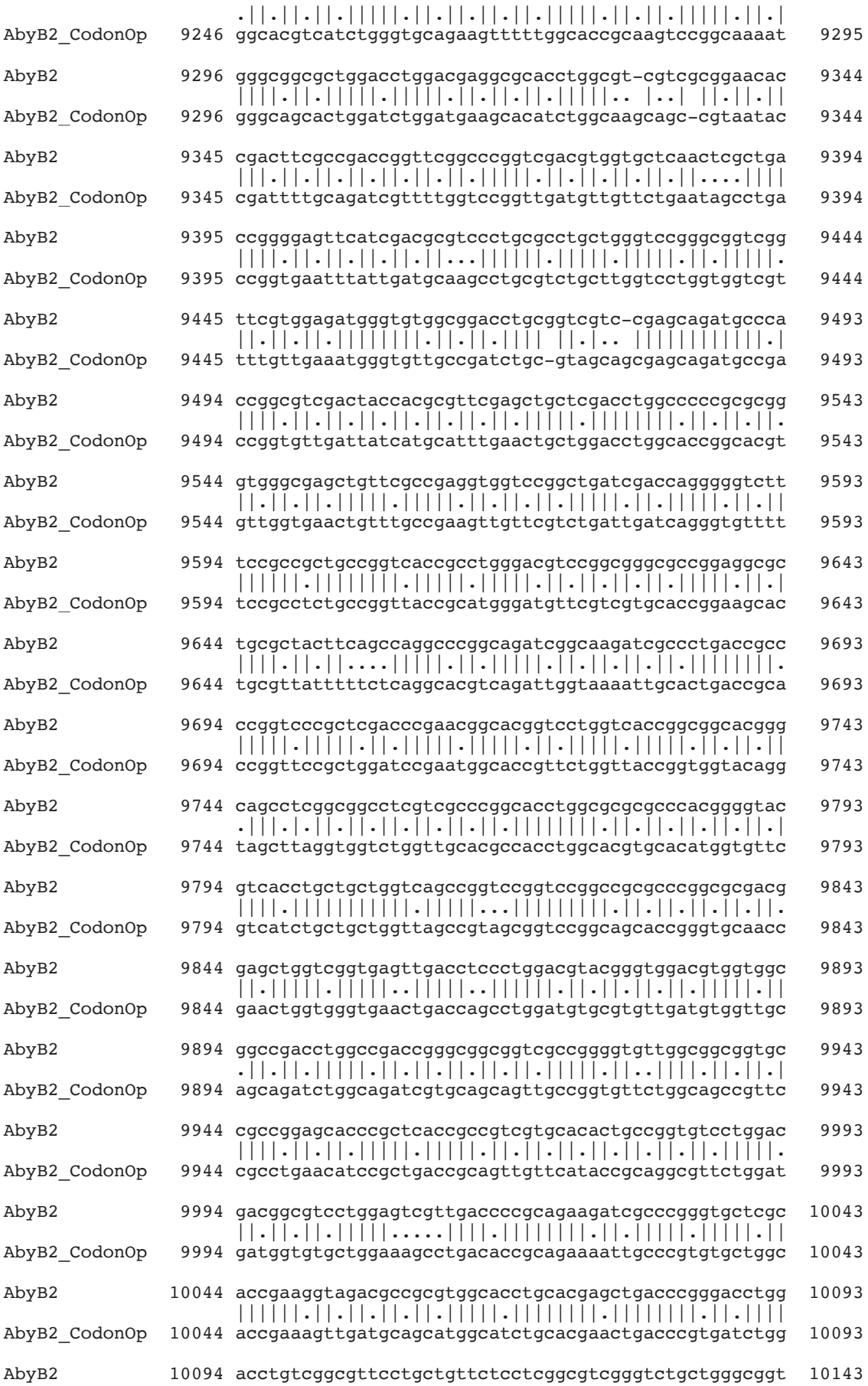
```
                    |.||....||.||.||.||.||.|||||.||.|||||.||.||.|||
AbyB2_CodonOp   4997 ttgttagcgagctggatgcagttggtccgtttcctaccgatcgtggttgg   5046

AbyB2           5047 ccggcggacctgtacgaccccgacccggaggccaccggccgcacgtacgc   5096
                     ||.||.||.|||||.||.||.||.||.||.|||||.||.||.||.||
AbyB2_CodonOp   5047 cctgcagatctgtatgatcctgatcctgaagcaaccggtcgtacctatgc   5096

AbyB2           5097 gcgtgagggcggcttcctctacgacgccgccgggttcgacccggaattct   5146
                     .|||||.||.||.||.||.||.|||||.||.||.||.||.|||||||.|
AbyB2_CodonOp   5097 acgtgaaggtggttttctgtatgacgcagcaggttttgatccggaatttt   5146

AbyB2           5147 tcggcatcagcccccgcgaggccaccggcatggacccgcagcagcggctg   5196
                     |.||.||.||.||.||.||.||||||||.|||||.||||||||||||.|||
AbyB2_CodonOp   5147 ttggtattagtccgcgtgaagccaccggtatggatccgcagcagcgtctg   5196

AbyB2           5197 ctgctgcagaccggctgggaggtgttcgagcgggccgggatcgaccccac   5246
                     .|.||||||||||.|||||.||.||.||.||.||.||.||.||.||.||
AbyB2_CodonOp   5197 ttactgcagaccggttgggaagtttttgaacgtgcgggtattgatccgac   5246

AbyB2           5247 ggcgctgcgcggcagccgtaccggggtcttcgcgggggtcgtctacaccg   5296
                     .||..|.||.||.||||||||||.||.||.|||||.||.||.||.||||
AbyB2_CodonOp   5247 cgcattacgtggtagccgtaccggtgtgtttgcgggtgttgtttataccg   5296

AbyB2           5297 actacggctcccgggccgacccgatccccgccgacctggagggatacctc   5346
                     |.||.||.||.||.|||||.|||||.||.||.||.|||||.||.||..|.
AbyB2_CodonOp   5297 attatggttcacgtgccgatccgattccggcagatctggaaggttattta   5346

AbyB2           5347 ggcatcggcagcgccggcagcatcgcctccggccgcatcgcctacaccct   5396
                     ||.||.||.||.||.||.|||||.||...|||.||.||.||.||.|||||
AbyB2_CodonOp   5347 ggtattggtagtgcaggtagcattgcgagcggtcgtattgcatataccct   5396

AbyB2           5397 cggcctggagggcccggcggtcaccgtggacaccgcgtgct-cctcgtcg   5445
                     .||..|.||.||.|||||.||.|||||.||.|||||.||.. |..|..|
AbyB2_CodonOp   5397 gggtttagaaggtccggcagttaccgttgataccgcatgtagcagcagc-   5445

AbyB2           5446 ctggtggcgctgcacctggccgtgcagtcgctgcgccgcggcgagtgcga   5495
                     |||||.||.|||||.|||||.||.|||...|||||.||.||.||.||.||
AbyB2_CodonOp   5446 ctggttgcactgcatctggcagttcagagcctgcgtcgtggtgaatgtga   5495

AbyB2           5496 cctggccctggcgggcggtgcgacggtgctgtccaacccggacatcttcg   5545
                     .|||||..|.||.||.|||||.||.||.||||..|||.|||||.||.||.|
AbyB2_CodonOp   5496 tctggcattagccggtggtgccaccgttctgagcaatccggatatttttg   5545

AbyB2           5546 tcggcttctcgcgccagcggggcctgtccccggacagccgttgcaaggcg   5595
                     |.||.||....||.|||||.||.||||...|||||.||||||||.||.||.
AbyB2_CodonOp   5546 ttggttttagccgtcagcgtggtctgagtccggatagccgttgtaaagca   5595

AbyB2           5596 ttcgccgccgccgccgacggcaccgccttcgccgagggtgtcggcctgct   5645
                     ||.||.||.||.|||||.||||||||.||.||||||.|||||.||.|||||
AbyB2_CodonOp   5596 tttgcagcagcagccgatggcaccgcatttgccgaaggtgttggtctgct   5645

AbyB2           5646 gctggtgcagcggctcgccgacgcgcggcgcgatggtcggccggtgctgg   5695
                     ||||||.|||||.||.||.||.||.||.||.||||||||.|||||.||||
AbyB2_CodonOp   5646 gctggttcagcgtctggcagatgcacgtcgtgatggtcgtccggttctgg   5695

AbyB2           5696 ccgtcatccggggcaccgccatcaaccaggacggcgcgtccaacgggctc   5745
                     |.||.||.||.||||||||||.||.|||||.||.||...|||.||.||.
AbyB2_CodonOp   5696 cagttattcgtggcaccgccattaatcaggatggtgcaagcaatggtctg   5745

AbyB2           5746 accgcacccaacgggccgtcgcagcagcgggtcatcctgggtgcgctggc   5795
                     ||||||||.||.||.|||...||||||||.||.||.||||||||.|||||
AbyB2_CodonOp   5746 accgcaccgaatggtccgagccagcagcgtgttattctgggtgcactggc   5795

AbyB2           5796 cgacgcggggctgcgcccgtccgacgtcgacgtggtggaggcgcacggca   5845
                     |||.||.||.|||||.|||...||.||.||.||.||.||.||.||.|||||
AbyB2_CodonOp   5796 cgatgcaggtctgcgtccgagtgatgttgatgttgttgaagcacatggca   5845

AbyB2           5846 ccggcaccaccctgggcgacccgatcgaggcgcaggcgatcatcgccacc   5895
```

```
                    ||||.||.||||||||.||.|||||.||.||.|||||.||.||.||.|||
AbyB2_CodonOp  5846 ccggtacaaccctgggtgatccgattgaagcccaggcaattattgcaacc  5895

AbyB2          5896 tacgggcagggccgcgacgagccgctgctgctcggctcgctgaagtcgaa  5945
                    ||.||.|||||.||.||.||.|||||||||.|.||....|||||....||
AbyB2_CodonOp  5896 tatggtcagggtcgtgatgaaccgctgctgttaggtagcctgaaaagcaa  5945

AbyB2          5946 catcggccacacccaggccgcagccggtgtcggcggcgtgatcaagatgg  5995
                    .||.||.||.||||||||.||.||.||.||.||.||.|||||.||.||||
AbyB2_CodonOp  5946 tattggtcatacccaggcagccgcaggcgttggtggtgtgattaaaatgg  5995

AbyB2          5996 tcgccgccatgcgacacgggctggtgccccggaccctgcacgtcgacgag  6045
                    |.||.||.|||||.||.||.|||||.||.||.||.|||||.||.||.||.
AbyB2_CodonOp  5996 ttgcagcaatgcgtcatggtctggttccgcgtacactgcatgtggatgaa  6045

AbyB2          6046 ccgaccccgcacgtggactggtcggcgggtgcggtccggctggtcaccga  6095
                    |||||.|||||.||.||.|||...||.||||||.||.||.|||||.|||||
AbyB2_CodonOp  6046 ccgacaccgcatgttgattggagtgccggtgcagttcgtctggttaccga  6095

AbyB2          6096 ggcgcggccctggccggagtcgaaccggccgcgccgcgccggggtgtcct  6145
                    .||.||.||.|||||.||....||.||.|||||.||.|||||.||...|.
AbyB2_CodonOp  6096 agcacgtccgtggccagaaagcaatcgtccgcgtcgtgccggtgttagca  6145

AbyB2          6146 cgttcggcatgagcggcacgaacgcccacgtcgtcgtcgagcagggcgac  6195
                    ..||.||.||||||||||.||.||.||.||.||.||.||||||||.
AbyB2_CodonOp  6146 gctttggtatgagcggcaccaatgcacatgttgttgtggaacagggcgat  6195

AbyB2          6196 ccgctcgaggtgccgccggtgcgcgccggccgcctcgtaccggtgccggt  6245
                    ||||||.||.||||||.||.||.||.||.||||||.||.|||||.|||||
AbyB2_CodonOp  6196 ccgctggaagttccgcctgttcgtgcaggtcgcctggttccggttccggt  6245

AbyB2          6246 gtccgccgccaacccggcggcgctgcgccgacaggccgcccgtctgctcc  6295
                    |..|||.||.||.||||||.||.|||||.||.|||||.||.|||||||||.|
AbyB2_CodonOp  6246 gagcgcagcaaatccggcagcactgcgtcgtcaggcagcacgtctgctgc  6295

AbyB2          6296 ccgccgtggccgaccggcatcccgccgacgtggcccgtaccctcgcggcc  6345
                    |.||.||.||.||.||.||||||.||.||.||.||.|||||||||.||.|||
AbyB2_CodonOp  6296 cagcagttgcagatcgtcatccggcagatgttgcacgtaccctggcagcc  6345

AbyB2          6346 cgcacctccctcgccacccgggccgtcgtgctcgcggacgacgccgacga  6395
                    ||.|||..|||.||.||||||.||.||.||.||.||||||.||.||.||.||
AbyB2_CodonOp  6346 cgtaccagcctggcaacccgtgcagttgttctggcggatgatgcagatga  6395

AbyB2          6396 gctggccgagggcctacgggcgctggacgacgcgacgttcaccggtccgg  6445
                    .|||||||.|||||.||.||.||||||.||.||.||.||.||||||||||
AbyB2_CodonOp  6396 actggccgaaggcctgcgtgcactggatgatgccacctttaccggtccgg  6445

AbyB2          6446 tgggcgacgccgacgagccgggcaaggtggtcttcgtcttccccgggcag  6495
                    |.||.||.|||||||.||||||.||.||.||.||.||.||.||.||.||.
AbyB2_CodonOp  6446 ttggtgatgccgacgaaccgggtaaagttgtttttgtttttcctggtcaa  6495

AbyB2          6496 ggcgggcagtggaccggcatggctctggacctgtaccgcgacgagcccac  6545
                    ||.||.||||||||||.|||||.|||||.|||||.||.||.|||||.||
AbyB2_CodonOp  6496 ggtggtcagtggaccggtatggcactggatctgtatcgtgatgagccgac  6545

AbyB2          6546 cttccgcgagtccctggacgcctgcgccgccgcgctggcaccgcacgtgg  6595
                    |||.||.||...|||||||.||.||.||.||.|||||||||.||.|||||
AbyB2_CodonOp  6546 ctttcgtgaaagcctggatgcatgtgcagcggcactggcacctcatgtgg  6595

AbyB2          6596 actgggcgttgctcgacgtgctcgccgacgaggaggccctgcggcgggtg  6645
                    |.||||||..||||.||.||.||.|||||||||.||.|||||||||.||.||.
AbyB2_CodonOp  6596 attgggcactgctggatgttctggccgacgaagaagccctgcgtcgcgtt  6645

AbyB2          6646 gacgtcgtgcagcccgcgctctgggcggtgatggtgtcggtggcccggct  6695
                    ||.||.||.||||||.||.||.||||||.||.||||||....||.||.||.||
AbyB2_CodonOp  6646 gatgtggttcagcctgcactgtgggcagttatggttagcgttgcacgcct  6695

AbyB2          6696 ctggcagcaccacggcgtcacgcccgacgcggtggtcggccactcccagg  6745
```









```
                    |.|||...||.||.|||||||...|...||....||||||.|.||.|||
AbyB2_CodonOp  10094 atctgagcgcatttctgctgtttagcagcgcaagcggtctgttaggtggt  10143

AbyB2          10144 gcggggcaggccaactacgcggcggccaatgcgttcctggacgcgctggc  10193
                     ||.||.|||||.||.||.||.||.||||||.||.||||||.||.|||||
AbyB2_CodonOp  10144 gcaggtcaggcaaattatgcagcagcaaatgcctttctggatgcactggc  10193

AbyB2          10194 cacggcgcggcgccgcgccggcctgcccgccgtgtcgctggcctggggca  10243
                     .||.||.||.||.||.|||||.|||||.||.||....|||||.|||||.|
AbyB2_CodonOp  10194 aaccgcacgtcgtcgtgccggtctgcctgcagttagcctggcatggggta  10243

AbyB2          10244 tgtgggcgcgggccaccggcctgaccgcccacctgggcggcacggacctg  10293
                     |||||||.||.||.|||||.|||||.||.||..|.||.||.|||||.|||
AbyB2_CodonOp  10244 tgtgggcacgtgcgaccggtctgacagcacatttaggtggtacggatctg  10293

AbyB2          10294 ggccgcatcgaacgcggcggcctgctgccgatgaccgacgagcagggcct  10343
                     ||.||.||.|||||.||.||.||||||||||||||||.||.||.||.||
AbyB2_CodonOp  10294 ggtcgtattgaacgtggtggtctgctgccgatgaccgatgaacaaggtct  10343

AbyB2          10344 ggctttgttcgacgccacctggacggccgaccgtccggtgctggtgccgg  10393
                     ||||..||||.||.||||||||||||.||.||.||||||||||||||.||||
AbyB2_CodonOp  10344 ggcactgtttgatgccacctggaccgcagatcgtccggtgctggttccgg  10393

AbyB2          10394 cgccgctgcgcctggaccggggccgcaccggatccggtgtggtgccggcc  10443
                     |.||||||||.|||||.||.||.||.|||||...||||||.||.|||||.
AbyB2_CodonOp  10394 caccgctgcgtctggatcgtggtcgtaccggtagcggtgttgttccggca  10443

AbyB2          10444 gtgctgcgcgcgctggtgcggccggtgcgccgggtggcccgctcggcggg  10493
                     ||.||||||.||.||||||.||.||||||.||.||.||.||.||....||.||
AbyB2_CodonOp  10444 gttctgcgtgcactggttcgtccggttcgtcgcgttgcacgtagcgcagg  10493

AbyB2          10494 gacggcgtcgccggactcgctgcgcgagcggctgctgccgctgtccccga  10543
                     .||.||.||.|||||....||||||.||||||.||||||||||||..||||||
AbyB2_CodonOp  10494 caccgcatcaccggatagtctgcgtgagcgcctgctgccgctgagcccga  10543

AbyB2          10544 cggagcgcacggccctgctggtggacctggtccgtacacaggtcgcggcg  10593
                     |.||.||.||.||.||||||||.||.||||||.||||||.||||||.||.||.
AbyB2_CodonOp  10544 ccgaacgtaccgcactgctggttgatctggttcgtacccaggttgccgca  10593

AbyB2          10594 gtcctgggccacaccgacaccgacgccgtggtggtggaccgggcgttcaa  10643
                     ||.||||||.||.||||||.||||||.||.||.||.||.||.||.||.||
AbyB2_CodonOp  10594 gttctgggtcataccgataccgatgcagttgttgttgatcgtgcatttaa  10643

AbyB2          10644 ggacagcggtttcgactcgttgacggcggtggagctgcgcaaccgggtgt  10693
                     .||.||||||||.||.||..||||.||.||.||.||||||.||.||.||..
AbyB2_CodonOp  10644 agatagcggttttgattcactgacagcagttgaactgcgtaatcgtgtta  10693

AbyB2          10694 cccgcgccaccgggctgcggctgccacccaccgtggtgttcgaccgcccc  10743
                     .|||.||.||.||.||||||.||||||.||.||||||.||||||.||.||.
AbyB2_CodonOp  10694 gccgtgcaacaggtctgcgcctgcctccgaccgttgtgtttgatcgtccg  10743

AbyB2          10744 acgccggcggagttggccgcgcaccttctcgaccagctcgtgccgcctgc  10793
                     ||.|||||.||..||||.||.||.||.||.||.||||||.||.||.||.||
AbyB2_CodonOp  10744 acaccggcagaactggcagcccatctgctggatcagctggttcctccggc  10793

AbyB2          10794 cgacgggccggccggcgcggcgacgcccgcccgcaagacccgaaagcaac  10843
                     .||.||.||.|||||.||.||.||.||.||.||.||.|||||.||.||.|
AbyB2_CodonOp  10794 agatggtcctgccggtgcagcaactccggcacgtaaaacccgtaaacagc  10843

AbyB2          10844 tcgactcggccacggtcgaggagatcttcgacttgatcgactcccagctc  10893
                     |.||....||.||.||.||.||.||.||..|||||||...||||.|.
AbyB2_CodonOp  10844 tggatagcgcaaccgttgaagaaatttttgatctgatcgatagccagtta  10893

AbyB2          10894 ggccgggggtctcgcagcgactatcaggaggtcgacgccgggtga  10938
                     ||.||.||....||.||.||.|||||.||.||.|||||.||.|.|
AbyB2_CodonOp  10894 ggtcgtggtagccgtagtgattatcaagaagttgacgcaggttaa  10938
```

### A.2.2 AbyB2 Amino Acid Sequence

```
   1 MTTPSKGTDS IADQQKLREY LRRVTDDLLR TRRRLTEVES ADREPVAIVS MACRFPGGVA
  61 SPEDLWQLVA SGTDAISGFP DDRGWPLDEL YDPDPEHPGT STTRQGGFLH DAADFDPEFF
 121 GISPREALTI DPQQRLLLET AWEAVERAGI APDSLRGSRT GVFAGVMYGD YGARLRPIPA
 181 GFEGYMGTGS AGSVATGRIA YTLGLEGPAV SVDTACSSSL VALHLAAQAL RRGECDLALA
 241 GGVTVIATPE LFVEFSRQRG LSPDGRCKAF AASADGTGWA EGVGLVLVER LADARRNGHP
 301 VLALLRGSAV NQDGRSSQLS APNGPAQRRV IRAALASAGL EPAEVDLVEA HGTGTRLGDP
 361 IEAQALLAEY GQGRTEPLWL GSLKSNIGHT QAAAGVGGVI KVVQAMRHGL LPATLHADEA
 421 TPHVDWSVGD VRLLTEARDW PARERPRRAA VSSFGISGTN AHVILEEGDP DGVADAPPDD
 481 VLARKPVPVV LSAHTASALG PQAARLRAHL DAHPDLTVAD VAHSLATTRT PLAERAVLVA
 541 ADLDELRTAL DAVADGDEPP VRGTAGHPGG VVFVFPGQGA QWAGMALDLY REDEVFRAAL
 601 DDCERALAPH VDWSLRAVLA DADALGRVDV VQPALWAVMV SLAALWQHHG VTPDAVVGHS
 661 QGEIAAACVA GALSLEQAAA VVALRARAIT AVAGRGAMAS VSVPAQQITE RWGDRITVAV
 721 TNSADATVVA GEPEAVAEVV AAYDAEGVRA RVLPVDYASH SAHVEPVREP ILDALRDLTP
 781 TEARVPFHST VTGAEFDTRG LTADYWYTNL RSTVRFDQAV TRLREQGHRI FVEISPHPVL
 841 TPVLGEGAFG TLRRDEGDRR RFITSLGAVH AVGVPVDWSA AIGPARRVPL PTYAFQRSRY
 901 WLDAPARTGD ASGVGVGPTD HPLLGGAVDV AGDGTLVLTG RLVPGADRAA AELRVGGVPV
 961 LSGTALLDLA LRAGELAGLG AVGEFSVETP LVLSATAGWL QVVVAPAGAD GDREIGVYAR
1021 PDHEAPWTRH GHGVLVPATA QDLPPHRPVP GAPLAPDEAV ERLAAAGVEL ASAPTAVSGD
1081 ADGYVADLAS PPEGRFTLHP ALLDAALLPA FGRGDARLPS RWRGVRLPQP DGQPTRASVR
1141 RRDGDSWAVS LTDSAGAPVV EIAEVVLGPV PVVSAAGHHD PLFTLEWAPV ARPRGAAATE
1201 PVPHELPGDA GPAALAVLRG GLAEPDGPRQ VVVARGAGTV TGLVRCAQLE EPGRVGLVEW
1261 DGRDADALRD AVAAGLPQVA VHGAELRTPR LAPLTAPGRP VELDGTALIV GEAGALRDAV
1321 LRTLARHGVD RLVVVDVDGT APADVGVPTE VLTGDPTDRA VLAEAVHRAA NLRTVVHAVQ
1381 PTADAPLAAL SPEDLTALVE RIVAPARHLH ELTAHLPLTR FVLSGSAAGV LGGIGQAAVA
1441 AATTGLGALA ARRRADGLPA QVVAWGPSAG TRRLGLVSLD DARLAALFAQ VLAHDVDVVA
1501 APLVRAGLRG QARAGTLPVA LRALVPALPG GATGLAARLA GASPAEGRRL LLDTIRTHVA
1561 GVLGHDDASG IDERRAFKDL GFDSLTAIEL RNRLNTALGR TLPATLIFDH PSPGALAEHL
1621 RDDLLGRAAV AAAPVAVASD EPIAIVAMGC RYPGGIADPE ALWQAVVSEL DAVGPFPTDR
1681 GWPADLYDPD PEATGRTYAR EGGFLYDAAG FDPEFFGISP REATGMDPQQ RLLLQTGWEV
1741 FERAGIDPTA LRGSRTGVFA GVVYTDYGSR ADPIPADLEG YLGIGSAGSI ASGRIAYTLG
1801 LEGPAVTVDT ACSSSLVALH LAVQSLRRGE CDLALAGGAT VLSNPDIFVG FSRQRGLSPD
```

1861 SRCKAFAAAA DGTAFAEGVG LLLVQRLADA RRDGRPVLAV IRGTAINQDG ASNGLTAPNG

1921 PSQQRVILGA LADAGLRPSD VDVVEAHGTG TTLGDPIEAQ AIIATYGQGR DEPLLLGSLK

1981 SNIGHTQAAA GVGGVIKMVA AMRHGLVPRT LHVDEPTPHV DWSAGAVRLV TEARPWPESN

2041 RPRRAGVSSF GMSGTNAHVV VEQGDPLEVP PVRAGRLVPV PVSAANPAAL RRQAARLLPA

2101 VADRHPADVA RTLAARTSLA TRAVVLADDA DELAEGLRAL DDATFTGPVG DADEPGKVVF

2161 VFPGQGGQWT GMALDLYRDE PTFRESLDAC AAALAPHVDW ALLDVLADEE ALRRVDVVQP

2221 ALWAVMVSVA RLWQHHGVTP DAVVGHSQGE IAAAHVAGAL SLADAAAVVA LRARAITAIA

2281 GTGGMASVAL GVGEVTRRWG HTVAVAATNG PDTAVIAGDP GVLDHIAATC AAEEVRVKIL

2341 PVDYASHSAH VEALREELLA ALETVQPRAA EIAFCSTVTA EALDTTTLTA DYWYTNLRST

2401 VRYDETVRRL HAEGHRTFLE MSPHPVLTTV TEQVTGAVAL GTLRRDEGDR RRFLTALAEA

2461 YVTGVAVDWR PAVGADARLV DLPTYAFASD RYWLDATTRP VDATGLGLAA TAHPLLGAAV

2521 DLADDEGVLL TGRLSLDSHP WLADHTVAGV PLLPGAAFVE LCAQAAEAAG AAGVAELTLE

2581 TPCVLPERGG VDVQVQVRDG GLRVYSRSVG DAWVRNASGV LLPTEPPAPA GWGAWPPPGA

2641 QAVDVEGLYP QLAASGYGYG PAFRGLRAAW RRGEEVFAEV RLPEGLEPEG YGLHPALLDA

2701 ALHALAFGDF LGAGVRLPFA FTGVRVFATG ADILRVRLSP RGEDTVAVAL ADSTGAPVAE

2761 IESLVLRAAP SLEATAPHAP DVLVLDWTPL ALPDTPVTEP DLLVVAPSDA HDPVAATGRL

2821 VTSTIAELAG RLADDRGAVV VTRDAVAVRP GDPAADLAHA ALWGLLRSAQ TENPDRFTLV

2881 DTDGRPESAA VVAAAVATGE PQIAVREGRG YVPRLARAAA NRGLVPPPGA WRLEAAGTTV

2941 DELRLTEVTE APLPAGHVRV AVRACGLNFR DVLATLGVVP RDAPLGAEGA GVVVEVGVGV

3001 TGFAPGDRVY GFLQGAIGPR AVVDARLLAH LPAGWSFAQA ATAPAVCTTA YYALVTLADL

3061 RPGERVLIHS AAGGVGLAAG HLARHLGAEV FGTASPAKWA ALDLDEAHLA SSRNTDFADR

3121 FGPVDVVLNS LTGEFIDASL RLLGPGGRFV EMGVADLRSS EQMPTGVDYH AFELLDLAPA

3181 RVGELFAEVV RLIDQGVFPP LPVTAWDVRR APEALRYFSQ ARQIGKIALT APVPLDPNGT

3241 VLVTGGTGSL GGLVARHLAR AHGVRHLLLV SRSGPAAPGA TELVGELTSL DVRVDVVAAD

3301 LADRAAVAGV LAAVPPEHPL TAVVHTAGVL DDGVLESLTP QKIARVLAPK VDAAWHLHEL

3361 TRDLDLSAFL LFSSASGLLG GAGQANYAAA NAFLDALATA RRRAGLPAVS LAWGMWARAT

3421 GLTAHLGGTD LGRIERGGLL PMTDEQGLAL FDATWTADRP VLVPAPLRLD RGRTGSGVVP

3481 AVLRALVRPV RRVARSAGTA SPDSLRERLL PLSPTERTAL LVDLVRTQVA AVLGHTDTDA

3541 VVVDRAFKDS GFDSLTAVEL RNRVSRATGL RLPPTVVFDR PTPAELAAHL LDQLVPPADG

3601 PAGAATPARK TRKQLDSATV EEIFDLIDSQ LGRGSRSDYQ EVDAG

### A.2.3 AbyB3 Amino Acid Sequence

```
   1 MSETREEKLV EYLKWVTGEL QETKAQLARA RAEREPIAIV SAACRLPGDV HSPEDLWRVV
  61 VDGVDAIGDV PTDRGWAVHE VYADAPAHRP LGGFLSDAAG FDAAFFGIGP HEATAMDPQH
 121 RLLLESSWEA VERAGIDPTT LRGSATGVYA GLVSQNYAAY GTPPELNGHL MTGTATSVAS
 181 GRIAYLLGLR GPAVTLDTAC SSSLVALHLA AQALRRGECD LALAGGATVM ATPALLAEFV
 241 TQGGLSPDAR CKAFAAAADG TGFAEGVGVL VLERLADARR HHRRVLAVLR GSAVNQDGAS
 301 NGLTAPSGPA QEEVIRAALA DAGLRPSDVD HVEAHGTGTR LGDPIEAAAL LATYGQDRAE
 361 PLWLGSVKSN IGHTQTAAGV AGVIKVIEAL RHERLPRTLH VDEPTPHVDW AAGKVRLLTE
 421 EQPWPRGKRR RVAGVSSFGI SGTNAHVLIE EGDPEPPPTP PTPSAHPVAW LLGAKTDSAL
 481 RAQAARLRQR FAVASHDPLD VAVALATTRT AFDRRAAVVA ADHDGLLRGL DALAAGETTP
 541 GRAVRGPTAF LFSGQGSQRV GMGTELRRVF PAFRDAWREV ADEVDRHLDQ PLDRVLADED
 601 LLLRTEYAQP ALFTLEVALV RLLGGWGLRP DLLLGHSLGE LVAAHVAGVL DLPDAVALVA
 661 ARGAAMQAAP AEGAMVAIRA AADEVRASLA GREHEVSVAA VNGPRSTVVS GDAGAVQEVA
 721 AHWAATGHRT SRLRVSHAFH SPHLDGVLDG FRAVAAGVRH HPPSIPVVSN LTGTVVEAFT
 781 AEHWVRHVRQ EVRFAAGVSA LTSAGVRRFV EVGPDAVLAA LAGENAPGTP VVATLRRDES
 841 EALTVVRALA ASHVTGARVD WRAFHDERTA AVPLPTYPFE HRRYWVSPPT GPAPTAPPPV
 901 ADEPPRESTP HEERLLDLVR THAAAVLGHD TPESVGPDDN FVEIGLSSFT ALEVRNRLCE
 961 GTGLELSPLA LFEHPTPAAL AEHLRAVRAA RT
```

### A.2.4 AbyB3ΔACP Amino Acid Sequence

```
   1 VSETREEKLV EYLKWVTGEL QETKAQLARA RAEREPIAIV SAACRLPGDV HSPEDLWRVV
  61 VDGVDAIGDV PTDRGWAVHE VYADAPAHRP LGGFLSDAAG FDAAFFGIGP HEATAMDPQH
 121 RLLLESSWEA VERAGIDPTT LRGSATGVYA GLVSQNYAAY GTPPELNGHL MTGTATSVAS
 181 GRIAYLLGLR GPAVTLDTAC SSSLVALHLA AQALRRGECD LALAGGATVM ATPALLAEFV
 241 TQGGLSPDAR CKAFAAAADG TGFAEGVGVL VLERLADARR HHRRVLAVLR GSAVNQDGAS
 301 NGLTAPSGPA QEEVIRAALA DAGLRPSDVD HVEAHGTGTR LGDPIEAAAL LATYGQDRAE
 361 PLWLGSVKSN IGHTQTAAGV AGVIKVIEAL RHERLPRTLH VDEPTPHVDW AAGKVRLLTE
 421 EQPWPRGKRR RVAGVSSFGI SGTNAHVLIE EGDPEPPPTP PTPSAHPVAW LLGAKTDSAL
 481 RAQAARLRQR FAVASHDPLD VAVALATTRT AFDRRAAVVA ADHDGLLRGL DALAAGETTP
 541 GRAVRGPTAF LFSGQGSQRV GMGTELRRVF PAFRDAWREV ADEVDRHLDQ PLDRVLADED
 601 LLLRTEYAQP ALFTLEVALV RLLGGWGLRP DLLLGHSLGE LVAAHVAGVL DLPDAVALVA
 661 ARGAAMQAAP AEGAMVAIRA AADEVRASLA GREHEVSVAA VNGPRSTVVS GDAGAVQEVA
```

```
721 AHWAATGHRT SRLRVSHAFH SPHLDGVLDG FRAVAAGVRH HPPSIPVVSN LTGTVVEAFT
781 AEHWVRHVRQ EVRFAAGVSA LTSAGVRRFV EVGPDAVLAA LAGENAPGTP VVATLRRDES
841 EALTVVRALA ASHVTGARVD WRAFHDERTA AVPLPTYPFE HRRYWVSPPT GPAPTAPPPV
901 ADEPPRESTP HE
```

### A.2.5 AbyU Amino Acid Sequence

```
  1 MTERLETRPQ ALLIKVPTEI VVKVVDDVDV AAPAVGQVGK FDDELYDEAG AQIGTSSGNF
 61 RIEYVRPTDG GLLTYYQEDI TLSDGVIHAE GWADFNDVRT SKWVFYPATG VSGRYLGLTG
121 FRQWRMTGVR KSAEARILLG E
```

## A.3 Sequence Alignment of mPKS KS-AT Didomains

AbyB3ΔACP was aligned with the sequences of both didomains solved by x-ray crystallography from the DEBS polyketide synthases. A high level of sequence identity is observed, with AbyB3ΔACP displaying 48 % sequence identity to each of DEBS M3 and M5, which, in turn, display 57 % sequence identity to each other.

Small hydrophobic residues are coloured red, acidic residues coloured blue, basic amino acids are coloured in magenta and hydroxly, sulfhydryl, and amine residues are coloured green. The KS active site cysteine and the AT active site serine residues are highlighted in yellow.

```
AbyB3ΔACP       ---VSETREEKLVEYLKWVTGELQETKAQLA-RARAEREPIAIVSAACRLPGDVHSPEDL    56
DEBS_M3         ----------------------------------MELESDPIAIVSMACRLPGGVNTPQRL    27
DEBS_M5         MSGDNGMTEEKLRRYLKRTVTELDSVTARLREVEHRAGEPIAIVGMACRFPGDVDSPESF    60
                                                 .   :*****. ***:**.*.:*: :

AbyB3ΔACP       WRVVVDGVDAIGDVPTDRGWAVHEVYADAPA-----HRPLGGFLSDAAGFDAAFFGIGPH    111
DEBS_M3         WELLREGGETLSGFPTDRGWDLARLHHPDPDNPGTSYVDKGGFLDDAAGFDAEFFGVSPR    87
DEBS_M5         WEFVSGGGDAIAEAPADRGWEP------DPD------ARLGGMLAAAGDFDAGFFGISPR    108
                *..:  * :::.  *:****         *          **:*  *..*** ***:.*:
```

```
AbyB3ΔACP       EATAMDPQHRLLLESSWEAVERAGIDPTTLRGSATGVYAGLVSQNYAAYG--TPPELNGH   169
DEBS_M3         EAAAMDPQQRLLLETSWELVENAGIDPHSLRGTATGVFLGVAKFGYGEDTA-AAEDVEGY   146
DEBS_M5         EALAMDPQQRIMLEISWEALERAGHDPVSLRGSATGVFTGVGTVDYGPRPDEAPDEVLGY   168
                ** *****:*::** *** :*.** ** :***:****: *: . .*.     :  :: *:

AbyB3ΔACP       LMTGTATSVASGRIAYLLGLRGPAVTLDTACSSSLVALHLAAQALRRGECDLALAGGATV   229
DEBS_M3         SVTGVAPAVASGRISYTMGLEGPSISVDTACSSSLVALHLAVESLRKGESSMAVVGGAAV   206
DEBS_M5         VGTGTASSVASGRVAYCLGLEGPAMTVDTACSSGLTALHLAMESLRRDECGLALAGGVTV   228
                  **.* :******::* :**.**::::******.*.***** ::**:.*..:*:.**.:*

AbyB3ΔACP       MATPALLAEFVTQGGLSPDARCKAFAAAADGTGFAEGVGVLVLERLADARRHHRRVLAVL   289
DEBS_M3         MATPGVFVDFSRQRALAADGRSKAFGAGADGFGFSEGVTLVLLERLSEARRNGHEVLAVV   266
DEBS_M5         MSSPGAFTEFRSQGGLAADGRCKPFSKAADGFGLAEGAGVLVLQRLSAARREGRPVLAVL   288
                *::*. :.:*  * .*: *.*.* *. .*** *::**. :::*:**: ***. : ****:

AbyB3ΔACP       RGSAVNQDGASNGLTAPSGPAQEEVIRAALADAGLRPSDVDHVEAHGTGTRLGDPIEAAA   349
DEBS_M3         RGSALNQDGASNGLSAPSGPAQRRVIRQALESCGLEPGDVDAVEAHGTGTALGDPIEANA   326
DEBS_M5         RGSAVNQDGASNGLTAPSGPAQQRVIRRALENAGVRAGDVDYVEAHGTGTRLGDPIEVHA   348
                ****:*********:*******..*** ** ..*:. .*** ******** ******. *

AbyB3ΔACP       LLATYGQDR--AEPLWLGSVKSNIGHTQTAAGVAGVIKVIEALRHERLPRTLHVDEPTPH   407
DEBS_M3         LLDTYGRDRDADRPLWLGSVKSNIGHTQAAAGVTGLLKVVLALRNGELPATLHVEEPTPH   386
DEBS_M5         LLSTYGAERDPDDPLWIGSVKSNIGHTQAAAGVAGVMKAVLALRHGEMPRTLHFDEPSPQ   408
                ** *** :*    ***:***********:****:*::*.: ***: .:* ***.:**:*:

AbyB3ΔACP       VDWAAGKVRLLTEEQPWPRGKRRRVAGVSSFGISGTNAHVLIEEGDPEPPPTPPTPSAHP   467
DEBS_M3         VDWSSGGVALLAGNQPWRRGERTRRAAVSAFGISGTNAHVIVEEAPERERHRETTAHDGRP   446
DEBS_M5         IEWDLGAVSVVSQARSWPAGERPRRAGVSSFGISGTNAHVIVEEAPEADEPEP-APDSGP   467
```

```
                ::*  * * :::  : *  *:* * *.**:**********:**.        : .. *

AbyB3ΔACP       VAWLLGAKTDSALRAQAARLRQRFAVA-SHDPLDVAVALATTRTAFDRRAAVVAADHDGL    526
DEBS_M3         VPLVVSARSTAALRAQAAQIAELLE-RPDADLAGVGLGLATTRARHEHRAAVVASTREEA    505
DEBS_M5         VPLVLSGRDEQAMRAQAGRLADHLAREPRNSLRDTGFTLATRRSAWEHRAVVV-GDRDEA    526
                *  ::..:   *:****.:: : :      .  .... *** *:  ::**.** . ::

AbyB3ΔACP       LRGLDALAAGETTPGRAVR-------GPTAFLFSGQGSQRVGMGTELRRVFPAFRDAWRE    579
DEBS_M3         VRGLREIAAGAATADAVVEGVTEVDGRNVVFLFPGQGSQWAGMGAELLSSSPVFAGKIRA    565
DEBS_M5         LAGLRAVADGRIADRTATG--QARTRRGVAMVFPGQGAQWQGMARDLLRESQVFADSIRD    584
                : **  :* *  :   ..          ..::* ***:*  **. :*     .* .  *

AbyB3ΔACP       VADEVDRHLDQPLDRVLA---DEDLLLRTEYAQPALFTLEVALVRLLGGWGLRPDLLLGH    636
DEBS_M3         CDESMAPMQDWKVSDVLRQAPGAPGLDRVDVVQPVLFAVMVSLAELWRSYGVEPAAVVGH    625
DEBS_M5         CERALAPHVDWSLTDLLS---GARPLDRVDVVQPALFAVMVSLAALWRSHGVEPAAVVGH    641
                    :    *  :  :*    .   * *.: .**.**:: *:*. *  . *:.*  ::**

AbyB3ΔACP       SLGELVAAHVAGVLDLPDAVALVAARGAAMQAAPAEGAMVAIRAAADEVRASLAGREHEV    696
DEBS_M3         SQGEIAAAHVAGALTLEDAAKLVVGRSRLMRSLSGEGGMAAVALGEAAVRERLRPWQDRL    685
DEBS_M5         SQGEIAAAHVAGALTLEDAAKLVAVRSRVLRRLGGQGGMASFGLGTEQAAERIGRFAGAL    701
                * **:.******.* * **. **. *.  ::   .:*.*.:.  .   .   :      :

AbyB3ΔACP       SVAAVNGPRSTVVSGDAGAVQEVAAHWAATGHRTSRLRVSHAFHSPHLDGVLDGFRAVAA    756
DEBS_M3         SVAAVNGPRSVVVSGEPGALRAFSEDCAAEGIRVRDIDVDYASHSPQIERVREELLETTG    745
DEBS_M5         SIASVNGPRSVVVAGESGPLDELIAECEAEAHKARRIPVDYASHSPQVESLREELLTELA    761
                *:*:******.**:*: * :  .  .  * . :.  : *.:* ***::: : : :    .

AbyB3ΔACP       GVRHHPPSIPVVSNLTGTVVE--AFTAEHWVRHVRQEVRFAAGVSALTSAGVRRFVEVGP    814
```

```
DEBS_M3          DIAPRPARVTFHSTVESRSMDGTELDARYWYRNLRETVRFADAVTRLAESGYDAFIEVSP    805
DEBS_M5          GISPVSADVALYSTTTGQPIDTATMDTAYWYANLREQVRFQDATRQLAEAGFDAFVEVSP    821
                 .:      : . *.  .  ::   : : :*  ::*: ***  ..  *:.:*   *:**.*

AbyB3ΔACP        DAVLAALAGE-------NAPGTPVVATLRRDESEALTVVRALAASHVTGARVDWRAFHDE    867
DEBS_M3          HPVVVQAVEEAVEEADG-AEDAVVVGSLHRDGGDLSAFLRSMATAHVSGVDIRWDVALPG    864
DEBS_M5          HPVLTVGIEATLDSALPADAGACVVGTLRRDRGGLADFHTALGEAYAQGVEVDWSPAFAD    881
                 . *:.              .: **.:*:** .    .  ::. ::. *. : *

AbyB3ΔACP        RTAAVPLPTYPFEHRRYWVSPPTGPAPTAPPPVADEPPRESTPHE-     912
DEBS_M3          -AAPFALPTYPFQRKRYWLQPAAPAAASD------ELAYRSSSVDK     903
DEBS_M5          -ARPVELPVYPFQRQRYWLPIPTGGRARDEDD---DWRYQ------     917
                  :  . **.***::::***:   :           :   .
```